Document AN2 WORLD INTELLECT Appl. No. 09/937,484

International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C12N 15/31, 1/21, C12P 21/02, C07K 14/33, A61K 38/16, 39/08

A1

(11) International Publication Number:

WO 98/07864

(43) International Publication Date:

26 February 1998 (26.02.98)

(21) International Application Number:

PCT/GB97/02273

08/782,893 (CIP)

(22) International Filing Date:

(60) Parent Application or Grant (63) Related by Continuation

22 August 1997 (22.08.97)

(30) Priority Data:

US

Filed on

9617671.4 9625996.5

23 August 1996 (23.08.96)

GB 13 December 1996 (13.12.96) GB

27 December 1996 (27.12.96)

FOSTER, Keith, Alan [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB).

Conrad, Padraig [GB/GB]; Microbiological Research

Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB).

(74) Agent: SCHLICH, George, William; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).

(71) Applicants (for all designated States except US): MICRO-BIOLOGICAL RESEARCH AUTHORITY CAMR (CEN-TRE FOR APPLIED MICROBIOLOGY & RESEARCH) [GB/GB]; Porton Down, Salisbury, Wiltshire SP4 0JG (GB). THE SPEYWOOD LABORATORY LIMITED [GB/GB]: 14 Kensington Square, London W8 5HH (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SHONE, Clifford, Charles [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). QUINN, (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RECOMBINANT TOXIN FRAGMENTS

(57) Abstract

A polypeptide has first and second domains which enable the polypeptide to be translocated into a target cell or which increase the solubility of the polypeptide, or both, and further enable the polypeptide to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis. The polypeptide thus combines useful properties of a clostridial toxin, such as a botulinum or tetanus toxin, without the toxicity associated with the natural molecule. The polypeptide can also contain a third domain that targets it to a specific cell, rendering the polypeptide useful in inhibition of exocytosis in target cells. Fusion proteins comprising the polypeptide, nucleic acids encoding the polypeptide and methods of making the polypeptide are also provided. Controlled activation of the polypeptide is possible and the polypeptide can be incorporated into vaccines and toxin assays.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL.			•		-		apprications under the
AM AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM AT	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
	Austria	FR	France	LU	Luxembourg	SN	
NU.	Australia	GA	Gabon	L.V	Latvia	SZ.	Senegal
17.	Azerbaijan	GB	United Kingdom	MC	Мопасо	TD	Swaziland
A	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Chad
B	Barbados	GH	Ghana	MG	Madagascar		Togo
E	Belgium	GN	Guinea	MK	The former Yugoslav	TJ	Tajikistan
F	Burkina Faso	GR	Greece		Republic of Macedonia	TM	Turkmenistan
G	Bulgaria	HU	Hungary	ML	Mali	TR	Turkey
J	Benin	IE	Ireland	MN	Mongolia	TT	Trinidad and Tobago
R	Brazil	IL	Israel	MR	Mongona Mauritania	UA	Ukraine
Y	Belarus	15	Iceland	MW	Malawi	UG	Uganda
A	Canada	IT	Italy	MX	Mexico'	US	United States of America
F	Central African Republic	JP	Japan	NE		UZ	Uzbekistan
G	Солдо	KE	Kenya	NL	Niger Netherlands	VN	Viet Nam
Ħ	Switzerland	KG	Kyrgyzstan	NO		YU	Yugoslavia
i	Côte d'Ivoire	KP	Democratic People's	NZ NZ	Norway	ZW	Zimbabwe
M	Cameroon		Republic of Korea		New Zealand		
N	China	KR	Republic of Korea	PL	Poland		
U	Cuba	KZ.	Kazakstan	PT	Portugal		
Z	Czech Republic	LC	Saint Lucia	RO	Romania		
3	Germany	LI	Liechtenstein	RU	Russian Federation		
K	Denmark	LK	Sri Lanka	SD	Sudan		
E	Estonia	LR		SE	Sweden		
		LR	Liberia	SG	Singapore		

WO 98/07864 PCT/GB97/02273

RECOMBINANT TOXIN FRAGMENTS

This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the $H_{\rm C}$ domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the $H_{\rm N}$ domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The $H_{\rm N}$ domain also has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the H_c , amino acid residues 449-871 for the H_N and residues 1-448 for the LC. Digestion with trypsin effectively degrades the H_C domain of the BoNT/A to generate a non-toxic fragment designated LH_N ,

which is no longer able to bind to and enter neurons (Fig. 1). The LH_N fragment so produced also has the property of enhanced solubility compared to both the parent holotoxin and the isolated LC.

It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

- (A) clostridial neurotoxin light chain:
- -a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin.
- (B) clostridial neurotoxin heavy chain H_N domain:
- -a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.
- -the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.
- -the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.
- -the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.
- (C) clostridial neurotoxin heavy chain H_c domain.
- -a portion of the heavy chain which is responsible for binding of the native

holotoxin to cell surface receptor(s) involved in the intoxicating action of clostridial toxin prior to internalisation of the toxin into the cell.

The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

However, due to its extreme toxicity, the handling of native toxin is hazardous.

The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

Accordingly, the invention provides a polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to neuronal exocytosis and wherein said second domain is adapted (i) to translocate the polypeptide into the cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into the cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of any clostridial neurotoxin precursor that can be converted into toxin by proteolytic action. Accordingly, the invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the H_N of a clostridial toxin heavy chain, whilst lacking the functional aspects of a clostridial toxin H_C domain.

For the purposes of the invention, the functional property or properties of the H_N of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter to a H_N domain or to the functions of a H_N domain are references to this property or properties. The second domain is not required to exhibit other properties of the H_N domain of a clostridial toxin heavy chain.

A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin-this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH₄₂₃/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is of use if solubility is imparted to a composition made up of that polypeptide and one or more other components by one or more of said other components.

The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has, indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from Aplysia (sea slug), Drosophila (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

The polypeptide of the invention may be obtained by expression of a recombinant nucleic acid, preferably a DNA, and is a single polypeptide, that is to say not

cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

In a polypeptide according to the invention, said first domain preferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasma-membrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in J. Biol. Chem., Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment (ii) a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between the LC and H_N components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

In an embodiment of the invention described in an example below, the toxin light

chain and the portion of the toxin heavy chain are of botulinum toxin type A. In a further embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain H_N portion or a fragment or variant of a clostridial toxin heavy chain H_N portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the H_N domain. Teachings of regions within the H_N responsible for its function are provided for example in Biochemistry 1995, 34, pages 15175-15181 and Eur. J. Biochem, 1989, 185, pages 197-203. The variant has a different sequence from the H_N domain or fragment, though it too retains the function of the H_N domain. It is conveniently obtained by insertion, deletion and/or substitution of a H_N domain or fragment thereof. In embodiments of the invention, described below, it comprises (i) an N-terminal extension to a H_N domain or fragment, (ii) a C-terminal extension to a H_N domain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

These polypeptides of the invention are thus not composed of two or more polypeptides, linked for example by di-sulphide bridges into composite molecules. Instead, these polypeptides are single chains and are not active or their activity is

significantly reduced in an in vitro assay of neurotoxin endopeptidase activity.

Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated $H_{\rm C}$ of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragment of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated $H_{\rm C}$ of a clostridial toxin heavy chain.

In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated H_N of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and H_N sequences of botulinum toxin types A, B, C₁, D, E, F and G.

The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic IgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then determines the target for a polypeptide - immunoglobulin complex. Alternatively, the

polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

As noted above, by proteolytic treatment, for example using trypsin, of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide. A third aspect of the invention provides a composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the clostridial toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*. The activity of the derivative preferably approaches that of natural toxin, and is thus preferably at least 30% and most preferably at least 60% of natural toxin. The overall endopeptidase activity of the composition will, of course, also be determined by the amount of the derivative that is present.

While it is known to treat naturally produced clostridial toxin to remove the H_C domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

The invention enables production of the polypeptides and fusion proteins of the invention by recombinant means.

A fourth aspect of the invention provides a nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described above.

In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain. This recombinant product is designated LH₄₂₃/A (SEQ ID NO: 2).

In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polypeptide or a fusion protein and comprises nucleotides encoding residues 1-

1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated LH₇₂₈/B (SEQ ID NO: 20).

The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host.

The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, the fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

The LH_N/A derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal 1/2 of the heavy chain, the H_C domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recovered from *E. coli*, and from other recombinant expression hosts, is an inactive single chain peptide or if expression

hosts produce a processed, active polypeptide it is not a toxin. Endopeptidase activity of LH₄₂₃/A, as assessed by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H_C domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EP-A-0763131.

A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant LH_N derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus

of the LC at the putative trypsin sensitive region and also at the extreme C-terminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

The LH_N enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin serotype or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a / to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

Following this nomenclature,

LH ₄₂₃ /A	= SEQ ID NO: 2, containing the entire L-chain and 423
	amino acids of the H-chain of botulinum neurotoxin type
	A ;

₂LH₄₂₃/A = a variant of this molecule, containing a two amino acid extension to the N-terminus of the L-chain;

 $_2L_{/2}H_{423}/A$ = a further variant in which the molecule contains a two amino acid extension on the N-terminus of both the L-chain and the H-chain;

²L_{FXa/2}H₄₂₃/A = a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the H-chain component; and

 $_2L_{FXa/2}H_{423}/A$ -IGF-1 = a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gene for an embodiment of the invention designated LH₄₂₃/A;

- Fig. 3 is a graph comparing activity of native toxin, trypsin generated "native" LH_N/A and an embodiment of the invention designated $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) in an *in vitro* peptide cleavage assay;
- Fig. 4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;
- Fig. 5 shows the transition region of an embodiment of the invention designated L/₄H₄₂₃/A illustrating insertion of four amino acids at the N-terminus of the H_N sequence; amino acids coded for by the *Eco* 47 III restriction endonuclease cleavage site are marked and the H_N sequence then begins ALN...;
- Fig. 6 shows the transition region of an embodiment of the invention designated L_{FXa/3}H₄₂₃/A illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H-sequence; the N-terminal amino acid of the cleavage-activated H_N will be cysteine;
- Fig. 7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ -IGF-1, a fusion protein; the IGF-1 sequence begins at position G_{882} ;
- Fig. 8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ -CtxA14, a fusion protein; the C-terminal CtxA sequence begins at position Ω_{882} ;
- Fig.9 shows the C-terminal portion of the amino acid sequence of an

embodiment of the invention designated $L_{FXa/3}H_{423}/A-ZZ$, a fusion protein; the C-terminal ZZ sequence begins at position A_{890} immediately after a genenase recognition site (underlined);

show schematic representations of manipulations of

Figs. 10 & 11 polypeptides of the invention; Fig. 10 shows LH₄₂₃/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an Ig binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage sites R1, R2 and R3 and a C-terminal fusion peptide sequence;

Fig. 12 shows the trypsin sensitive activation region of a polypeptide of the invention;

shows Western blot analysis of recombinant LH₁₀₇/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH₁₀₇/B; panel B was probed with anti-T7 peptide tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lane 4 immunopurified LH₁₀₇/B.

The sequence listing that accompanies this application contains the following sequences:-

SEQ ID NO:

Sequence

1

DNA coding for LH₄₂₃/A

2	LH ₄₂₃ /A
3	DNA coding for 23LH423/A (Q2E,N26K,A27Y), of which an
	N-terminal portion is shown in Fig. 4.
4	$_{23}LH_{423}/A (O_2E, N_{26}K, A_{27}Y)$
5	DNA coding for 2LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y), of which an N-
	terminal portion is shown in Fig.4
6	₂ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
·	
7	DNA coding for native BoNT/A according to Binz et al
8	native BoNT/A according to Binz et al
9	DNA coding for L _{/4} H ₄₂₃ /A
10	L _{/4} H ₄₂₃ /A
11	DNA coding for L _{FX3} / ₃ H ₄₂₃ /A
12	L _{FXa} / ₃ H ₄₂₃ /A
13	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A-IGF-1
14	L _{FXa} / ₃ H ₄₂₃ /A-IGF-1
15	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A-CtxA14
16	$L_{FXa}/_3H_{423}/A-CtxA14$
17	DNA coding for L _{FXa/3} H ₄₂₃ /A-ZZ
18	L _{FXa/3} H ₄₂₃ /A-ZZ
19	DNA coding for LH ₇₂₈ /B
20	LH ₇₂₈ /B
21	DNA coding for LH ₄₁₇ /B
22	LH ₄₁₇ /B
23	DNA coding for LH ₁₀₇ /B
24	LH ₁₀₇ /B
25	DNA coding for LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
26	LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
27	DNA coding for LH ₄₁₇ /B wherein the first 274 bases are

28

modified to have an *E. coli* codon bias

DNA coding for LH₄₁₇/B wherein bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide

Example 1

A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain (H_c) of botulinum neurotoxin type A. This recombinant product is designated the LH₄₂₃/A fragment (SEQ ID NO: 2).

Construction of the recombinant product

The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an E. coli codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique Kpnl restriction site. The remainder of the LH_{423}/A coding sequence was PCR amplified from total chromosomal DNA from $Clostridium\ botulinum\$ and annealed to the synthetic portion of the gene.

The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polymerase amplified DNA (bases 914-1138 and 1976-2616) and the remainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The

assembled gene was then fully sequenced and cloned into a variety of *E.coli* plasmid vectors for expression analysis.

Expression of the recombinant gene and recovery of protein product

The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZZ18, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

Currently, E. coli harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth phase. Expression of the gene is then induced by addition isopropylthio- β -D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced temperature of 25°C. The cells are then harvested by centrifugation, are resuspended in a buffer solution containing 10 mM Na₂HPO₄, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH₄₂₃/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art.

The recombinant GST- LH_{423}/A is purified by adsorption onto a commercially prepared affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant LH_{423}/A is recovered in the non-adsorbed material.

Construct variants

A variant of the molecule, LH_{423}/A ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of LH_{423}/A producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A .

Two further variants of the gene sequence that have been expressed and the corresponding products purified are $_{23}LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

In yet another variant a gene has been produced which contains a Eco 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in the gene representing the interface of the heavy and light chains in native neurotoxin, and provides the capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the Eco 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, $L_{/4}H_{423}/A$ (SEQ ID NO: 10), which contains an additional four amino acids between amino acids 448 and 449 of L_{H423}/A at a position equivalent to the amino terminus of the

heavy chain of native BoNT/A.

A variant of the gene has been expressed, L_{FXa/3}H₄₂₃/A (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of L_{/4}H₄₂₃/A. The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated, and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention.

Variants of $L_{FXa/3}H_{423}/A$ have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.

Specific examples described are:

- (1) $L_{FXa/3}H_{423}/A$ -IGF-1 (SEQ ID NO: 14), in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity;
- (2) $L_{FXa/3}H_{423}/A$ -CtxA14 (SEQ ID NO: 16), in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and
- (3) $L_{\rm FXa/3}H_{\rm 4Z3}/A$ -ZZ (SEQ ID NO: 18), in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cleavage site located between the end of the clostridial heavy chain sequence and the sequence coding for the binding

ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a generase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing H_N function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.

It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.

Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.

The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.

Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single polypeptide

incorporating a domain equivalent to the light chain of botulinum neurotoxin type A and a domain equivalent to the H_N domain of the heavy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely GST, and a C-terminal extension providing a ligand binding domain, namely an IgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

Assay of product activity

The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains.

This activity is dependent on proteolytic modification of the recombinant GST- $_2$ LH $_{423}$ /A (Q_2 E, N_{26} K, A_{27} Y) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600 μ g/ml) is incubated at 37°C for 10-50 minutes with trypsin (10 μ g/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na $_2$ HPO $_4$, 1.8 mM KH $_2$ PO $_4$, pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

₂LH₄₂₃/A is more stable in the presence of trypsin and more active in the in vitro

peptide cleavage assay than is ₂₃LH₄₂₃/A. Both variants, however, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the recombinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moieties as would be obvious to those skilled in the art.

Example 2

As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

The gene sequences relating to this example were all assembled and expressed using methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

A gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide, LH₇₂₈/B (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

A gene has also been assembled coding for a variant polypeptide, LH_{417}/B (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus

equivalent by amino acid homology to that at the carboxy-terminus of the heavy chain fragment in native $\ensuremath{\mathsf{LH}_{\mathsf{N}}/\mathsf{A}}$.

A gene has also been assembled coding for a variant polypeptide, LH_{107}/B (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

Construct Variants

A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E.coli* codon usage bias.

For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B were synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. In addition, terminal restriction endonuclease sites of the synthetic products were constructed to facilitate insertion of these products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27). Similarly the sequence could be inserted into other genes of the examples.

Another variant sequence equivalent to nucleotides 691 to 1641 of SEQ ID NO: 21

, and employing non-native codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

Example 3

An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-2LH423/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA11) reactive against a conformation dependent epitope on the native LH_N/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na₂HPO₄ 1.15 g/l, KH₂PO₄ 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

For immunisation, 20 μ g of GST-₂LH₄₂₃/A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to its derivative LH_N/A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na₂HPO₄ anhydrous 10 g/l, gelatin (Difco) 2 g/l, pH 6.5-6.6) containing a dilution range from 0.5 μ g (5X10-6 g) to 5 picograms (5X10-12 g). Aliquots of 0.5 ml were injected into mice intraperitoneally and deaths recorded

over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti- $GST_{-2}LH_{423}/A$ antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1. Neutralisation of botulinum neurotoxin in mice by guinea pig anti-GST-2LH423/A antiserum.

Botulinum Toxin/mouse													
Survivors On Day	0.5µg	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)						
1	0	4	4	4	4	4	4						
2	•	4	4	4	4	4	4						
3	-	4	4 .	4	4	4	4						
4	-	4.	4	4	4	4	4						

TABLE 2. Neutralisation of botulinum neurotoxin in mice by non-immune guinea pig antiserum.

Botulinum Toxin/mouse													
Survivors On Day	0.5µg	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)						
1	0	o	0	0	o	2	4						
2	-	-	•	-	-	o	4						
3 .	-	-	•	-	-	-	4						
4	-	-	•	-	-		4						

Example 4

Expression of recombinant LH₁₀₇/B in E. coli.

As an exemplification of the expression of a nucleic acid coding for a LH_N of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide LH_{107}/B (SEQ ID

NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in $E.\ coli$ BL21 (DE3) (New England BioLabs, Beverley, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for LH_{423}/A .

Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverley, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7 antibody and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT:
 - (A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: THE SPEYWOOD LABORATORY LIMITED
 - (B) STREET: 14 Kensington Square

 - (C) CITY: London (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): W8 5HH
 - (A) NAME: FOSTER; Keith Alan
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: QUINN; Conrad Padraig
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: SHONE; Clifford Charles
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 (D) STATE: Wiltshire

 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
- (ii) TITLE OF INVENTION: Recombinant Toxin Fragments
- (iii) NUMBER OF SEQUENCES: 28
- (iv) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
- (2) INFORMATION FOR SEQ ID NO: 1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2616 base pairs

 - (B) TYPE: nucleic acid(C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

							-										
ATO Met	G CAC	TTO Phe	C GTO	G AAG L Asi	C AAC n Lys	G CAC	TTO Phe	C AA(⊇ Asr	TAT	rys	G GAG G Asi	C CC1	GT#	A AAG Ası	GGT Gly	•	48
GT1 Val	GAC Asp	ATT	GCC Ala	י דעי	C ATO	Lys	ATT	CCA Pro 25) Asn	GCC Ala	GGC Gly	CAG Gln	ATC Met	Glr	CCG		96
GTG Val	AAG Lys	GCT Ala 35		AAC Lys	G ATT	CAT His	AAC Asn 40	rpys	ATC	TGG	GTT Val	T ATT	Pro	GAA Glu	CGC		144
GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	Pro	CCG Pro	CCG	GAA Glu	٠	192
GCA Ala 65	2,2	CAG Gln	GTG Val	CCA Pro	GTT Val 70	TCA Seŕ	TAC Tyr	TAC	GAT Asp	TCA Ser 75	ACC	TAT	CTG Leu	AGC Ser	ACA Thr 80		240
	71011	Q1u	Буз	85	ASII	lyr	Leu	Lys	90	Val	Thr	AAA Lys	Leu	Phe 95	Glu		288
			100	1111	vsħ	Leu	GIŸ	105	met	Leu	Leu	ACC Thr	Ser 110	Ile	Val		336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys		384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr		432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160		480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	•	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	ser	ACT Thr	GIn	Tyr	Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe		576
ACG Thr	FILE	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu		624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	Val	ACC Thr 220	CTG (Leu'.	GCA Ala :	CAC His	GAG Glu		672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	Gly	ATT (Ile . 235	GCG Ala	ATT . Ile .	AAC Asn	Pro	AAC Asn 240		720

CGG	GTC Val	TTC Phe	AAC Lys	GT1 Val 245	. ASI	ACC Thr	AA(C GC	C TA a Ty 25	r Ty	C GA r Gl	G A1 u Me	TG AC	er G	GT : ly :	ITA Leu		768
GA/ Gli	A GTA ı Val	AGC Ser	Phe 260	GIU	GAA Glu	CTG Leu	Arg	265	r Phe	C GG e Gly	r GG y Gl	C CA y Hi	T GA S As 27	p A	CG A	AAG Jys		816
TT1 Phe	ATC Ile	GAC Asp 275	ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	1 GI	G TT(C CG1	r CT	G TA u Ty 28	r Ty	C TA	AC A	AC sn		864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTC Leu	AAC 1 Asn	AAC Lys	GC' Ala 300	a Ly	G TC s Se	C AT	T G	TG		912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	Va]	T TT	F AA.	A GA s Gl	u L	AA ys 20	·	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	Phe	TCC	GT/	A GAT	r AA D Ly 33	s L	TA eu		1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	-AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	AC/ Th: 350	Gl	G G	AT sp		1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr	Tyr	TT(G AA	AT sn		1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AA7 Asr	T TA	C T		1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GIA	TTT Phe 390	AAT ' Asn '	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AA As	n	:	1200
TTT Phe	AAT Asn	GGT (GIn A	AAT Asn 405	ACA Thr	GAA 2 Glu 1	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	Le	A u	1	L248
AAA Lys	AAT ' Asn '	rne :	ACT (Thr (420	GGA :	TTG ' Leu i	TTT (Phe (ilu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	Val	AG. Ar	A g	1	.296
GGG A	ile .	ATA 1 Ile 1 135	ACT 1 Thr S	CT / Ser I	AAA 1 Lys :	Thr I	LAA Lys :	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AA(Ly:	3	1	.344
GCA :	TTA / Leu / 150	AAT (Asn <i>A</i>	GAT T Asp I	TTA 7 Leu (Cys]	ATC A (le L 155	AA (GTT .	AAT . Asn .	Asn '	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TT: Phe	r •	1	392
AGT (Ser I 465	CCT T	CA C	GAA G Glu A	sp A	AT I Asn E	TTT A Phe I	CT I	AAT (Asn)	Asp 1	CTA. Leu 1 475	AAT Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu	ı	1	440
ATT A	ACA T	CT G	sp T	CT A hr A 85	AT A sn I	ATA G	AA (Ala A	GCA (Ala (490	GAA (Glu (GAA Glu	AAT Asn	ATT Ile	AGT Ser 495	TTA Leu	l.	1	48 8
GAT T Asp I	TA A Leu I	le G	AA C ln G	AA T	AT I	AT T	eu 1	ACC Thr I	TTT I Phe I	AAT 1 Asn E	TTT Phe	Asp .	AAT Asn 510	GAA Glu	CCT Pro	1	1	536

-	32	-
---	----	---

														CAA Gln		15	84
														TAT Tyr	GAG Glu	16	32
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT His	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	16	80
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu	17	28
														GTA Val	AAG Lys	. 17	76
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	Trp	GTA Val	GAA Glu	18	24
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT A sp	TTT Phe	ACC Thr 615	Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	18	72
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	19	20
TTA Leu	AAT Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	GAT Asp 650	GAT Asp	TTT Phe	GTA Val	GGT Gly	GCT Ala 655	TTA Leu	19	68
ATA Ile	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT Ile	GCA Ala	20	16
ATA Ile	CCT Pro	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys	20	64
GTT Val	CTA Leu 690	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	AAT Asn	GAA Glu	21	12
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720	21	60
GTT Val	AAT Asn	ACA Thr	CAG Gln	ATT Ile 725	GAT Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu	22	80
GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	22	56
CAA Gln	TAT Tyr	ACT Thr 755	Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	Asn	ATT	AAT Asn	TTT Phe	AAT Asn 765	ATT	GAT Asp	GAT Asp	23	04
TTA Leu	AGT Ser 770	Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	Met	ATT	AAT Asn	ATA Ile	23	152

AAT Asn 785	гÀг	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2,400
ATC Ile	CCT Pro	TAT	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
GAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2592
					ATT Ile 870		TAA *									2616

- (2) INFORMATION FOR SEQ ID NO: 2:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 872 amino acids
 - (B) TYPE: amino acid (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 170

Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe
Thr	Phe	Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp		Asn .205	Pro	Leu	Leu
Gly	Ala 210	Gly	Lys	Phe	Ala	Thr 215	Asp	Pro	Ala	Val	Thr 220	Leu	Ala	His	Glu
Leu 225	Ile	His	Ala	Gly	His 230	Arg	Leu	Tyr	Gly	Ile 235	Ala	Ile	Asn	Pro	Asn 240
Arg	Val	Phe	Lys	Val 245	Asn	Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	Leu
Glu	Val	Ser	Phe 260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys
Phe	Ile	Asp 275	Ser	Leu	Gln	Glu	Asn 280	Glu	Phe	Arg	Leu	Tyr 285	Tyr	Tyr	Asn
Lys	Phe 290	Lys	Asp	Ile	Ala	Ser 295		Leu	Asn	Lys	Ala 300	Lys	Ser	Ile	Val
Gly 305	Thr	Thr	Ala	Ser	Leu 310	Gln	Tyr	Met	Lys	Asn 315	Val	Phe	Lys	Glu	Lys 320
Tyr	Leu	Leu	Ser	Glu 325	Asp	Thr	Ser	Gly	Lys 330	Phe	Ser	Val	Asp	Lys 335	Leu
Lys	Phe	Asp	Lys 340	Leu	Tyr	Lys	Met	Leu 345	Thr	Glu	Ile	Tyr	Thr 350	Glu	Asp
Asn	Phe	Val 355	Lys	Phe	Phe	Lys	Val 360	Leu	Asn	Arg	Lys	Thr 365	Tyr	Leu	naA
Phe	Asp 370	Lys	Ala	Val	Phe	Lys 375	Ile	Asn	Ile	Val	Pro 380	Lys	Val	Asn	Tyr
Thr 385	Ile	Tyr	Asp	Gly	Phe 390	Asn	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	Asn 400
Phe	Asn	Gly	Gln	Asn 405	Thr	Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu
Lys	Asn	Phe	Thr 420	Gly	Leu	Phe	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	Arg
Gly	Ile	Ile 435	Thr	Ser	Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys
Ala	Leu 450	Asn	Asp	Leu	Cys	Ile 455	Lys	Val	Asn	Asn	Trp 460	Asp	Leu	Phe	Phe
Ser 465	Pro	Ser	Glu	Asp	Asn 470	Phe	Thr	Asn	Asp	Leu 475	Asn	Lys	Gly	Glu	Glu 480
Ile	Thr	Ser	Asp	Thr 485	Asn	Ile	Glu	Ala	Ala 490	Glu	Glu	Asn	Ile	Ser 495	Leu
Asp	Leu	Ile	Gln 500	Gln	Tyr	Tyr	Leu	Thr 505	Phe	Asn	Phe	Asp	Asn 510	Glu	Pro
Glu	Asn	Ile 515	Ser	Ile	Glu	Asn	Leu 520	Ser	Ser	Asp	Ile	Ile 525	Gly	Gln	Leu

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 555 560

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
580 585 590

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
595 600 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr
610 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655

Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 .685

Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700

Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
705 710 715 720

Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735

Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750

Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
755 760 765

Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 775 780

Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785 790 795 800

Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855

Thr Phe Thr Glu Tyr Ile Lys

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2685 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double

 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..2685
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGA Gly 1	TCC Ser	CCA Pro	GGA Gly	ATT Ile 5	CAT His	ATG Met	ACG Thr	TCG Ser	ACG Thr 10	CGT Arg	CTG Leu	CAG Gln	AAG Lys	CTT Leu 15	CTA Leu	48
GAA Glu	TTC Phe	GAG Glu	CTC Leu 20	CCG Pro	GGT Gly	ACC Thr	ATG Met	GAG Glu 25	TTC Phe	GTG Val	AAC Asn	AAG Lys	CAG Gln 30	TTC Phe	AAC Asn	96
TAT Tyr	AAG Lys	GAC Asp 35	CCT Pro	ĢTA Val	AAC Asn	GGT Gly	GTT Val 40	GAC Asp	ATT Ile	GCC Ala	TAC Tyr	ATC Ile 45	AAA Lys	ATT Ile	CCA Pro	144
		GGC Gly														192
ATC Ile 65	TGG Trp	GTT Val	ATT Ile	CCG Pro	GAA Glu 70	CGC Arg	GAT Asp	ACA Thr	TTT Phe	ACG Thr 75	AAC Asn	CCG Pro	GAA Glu	GAA Glu	GGA Gly 80	240
		AAC Asn														288
		ACC Thr														336
		ACC Thr 115														384
		CTG Leu														432
		GAC Asp														480
		CCA Pro														528
		CCC Pro														576
		GTG Val 195														624

ATT	CG?	y Phe	C AGO	CCA Pro	A GAC	TTC Phe 215	Thr	TT(GGT Gly	TTC / Phe	GAG Glu 220	ı Glı	G AG	C CT	G GAG u Glu	672
GT1 Val 225	. Asp	ACC Thi	AAC Asn	CCG Pro	CTG Leu 230	Leu	GGT Gly	GCA Ala	GGC Gly	Lys 235	Phe	C GC/	A ACT	r GA	CCA Pro 240	720
GCC Ala	GTG Val	ACC Thr	CTG Leu	GCA Ala 245	His	GAG Glu	CTG Leu	ATC	CAC His 250	Ala	GG1 Gly	CAT His	CG1	CTC Let 25	TAT Tyr	768
GGC Gly	ATT	Ala	Ile 260	Asn	CCG Pro	AAC Asn	CGC Arg	GTG Val 265	Phe	AAG Lys	GTT Val	AAC Asn	ACC Thr 270	Ası	GCC Ala	816
TAC Tyr	TAC	GAG Glu 275	Met	AGT Ser	GGT Gly	TTA Leu	GAA Glu 280	GTA Val	AGC Ser	TTC Phe	GAG Glu	GAA Glu 285	CTG Leu	CGC	ACG Thr	864
TTC Phe	GGT Gly 290	GGC Gly	CAT	GAT Asp	GCG Ala	AAG Lys 295	TTT Phe	ATC Ile	GAC Asp	AGC Ser	TTG Leu 300	CAG Gln	GAG Glu	AAC Asn	GAG Glu	912
TTC Phe 305	CGT Arg	CTG Leu	TAC Tyr	TAC Tyr	TAC Tyr 310	AAC Asn	AAG Lys	TTT Phe	AAA Lys	GAT Asp 315	ATT Ile	GCA Ala	AGT Ser	ACA Thr	CTG Leu 320	960
AAC Asn	AAG Lys	GCT Ala	AAG Lys	TCC Ser 325	ATT Ile	GTG Val	GGT Gly	ACC Thr	ACT Thr 330	GCT Ala	TCA Ser	TTA Leu	CAG Gln	TAT Tyr 335	ATG Met	1008
AAA Lys	AAT Asn	GTT Val	TTT Phe 340	AAA Lys	GAG Glu	AAA Lys	TAT Tyr	CTC Leu 345	CTA Leu	TCT Ser	GAA Glu	GAT Asp	ACA Thr 350	TCT Ser	GGA Gly	1056
AAA Lys	TTT Phe	TCG Ser 355	GTA Val	GAT Asp	AAA Lys	TTA Leu	AAA Lys 360	TTT Phe	GAT Asp	AAG Lys	TTA Leu	TAC Tyr 365	AAA Lys	ATG Met	TTA Leu	1104
ACA Thr	GAG Glu 370	ATT Ile	TAC Tyr	ACA Thr	GAG Glu	GAT Asp 375	AAT Asn	TTT Phe	GTT Val	AAG Lys	TTT Phe 380	TTT Phe	AAA Lys	GTA Val	CTT Leu	1152
Asn	Arg	Lys	Thr	Tyr	Leu	TAA neA	Phe	Asp	Lys	Ala	Val	TTT Phe	AAG Lys	ATA Ile	AAT Asn 400	1200
ATA Ile	GTA Val	CCT Pro	AAG Lys	GTA Val 405	AAT Asn	TAC Tyr	ACA Thr	ATA Ile	TAT Tyr 410	GAT Asp	GGA Gly	TTT Phe	AAT Asn	TTA Leu 415	AGA Arg	1248
AAT Asn	ACA Thr	AAT Asn	TTA Leu 420	GCA Ala	GCA Ala	AAC Asn	Phe .	AAT Asn 425	GGT Gly	CAA Gln	AAT Asn	ACA Thr	GAA Glu 430	ATT Ile	AAT Asn	1296
AAT Asn	ATG Met	AAT Asn 435	TTT Phe	ACT Thr	AAA Lys	CTA . Leu	AAA Lys 440	AAT Asn	TTT Phe	ACT Thr	Gly	TTG Leu 445	TTT Phe	GAA Glu	TTT Phe	1344
TAT Tyr	AAG Lys 450	TTG Leu	CTA Leu	TGT Cys	Val .	AGA (Arg (4 5 5	GGG :	ATA . Ile	ATA Ile	Thr	TCT Ser 460	A AA Lys	ACT Thr	AAA Lys	TCA Ser	1392
TTA Leu 465	GAT Asp	AAA Lys	GGA Gly	Tyr .	AAT Asn 470	AAG (Lys)	GCA 1	ITA . Leu .	Asn .	GAT Asp 475	TTA L	TGT . Cys	ATC . Ile	AAA Lys	GTT Val 480	1440

ААТ	' AAT	TGG	GAC	TTG	דייריים:	ىلىلىن .	י אריים	CCT	TO					_				
Asn	Asn	Trp	Asp	Leu 485	. Pne	Phe	Ser	Pro	Ser 490	Glu	A GAT	AAT Asr	TTT Phe	ACT Thr 495	AAT Asn	:	1488	-
GAT Asp	CTA Leu	AAT Asn	AAA Lys 500	GIA	GAA Glu	GAA Glu	ATT Ile	ACA Thr 505	Ser	GAT Asp	ACT Thr	AAT Asn	ATA Ile 510	Glu	GÇA Ala		1536	
GCA Ala	GAA Glu	GAA Glu 515	AAT Asn	ATT	AGT	TTA Leu	GAT Asp 520	TTA Leu	ATA Ile	CAA Gln	CAA Gln	TAT Tyr 525	Tyr	TTA Leu	ACC Thr		1584	
TTT Phe	AAT Asn 530	TTT Phe	GAT Asp	AAT Asn	GAA Glu	CCT Pro 535	GAA Glu	AAT Asn	ATT Ile	TCA Ser	ATA Ile 540	GAA Glu	AAT Asn	CTT Leu	TCA Ser		1632	
AGT Ser 545	GAC As p	ATT Ile	ATA Ile	GGC Gly	CAA Gln 550	TTA Leu	GAA Glu	CTT Leu	ATG Met	CCT Pro 555	AAT Asn	ATA Ile	GAA Glu	AGA Arg	TTT Phe 560	_	1680	
CCT Pro	AAT Asn	GGA Gly	AAA Lys	AAG Lys 565	TAT Tyr	GAG Glu	TTA Leu	GAT Asp	AAA Lys 570	TAT Tyr	ACT Thr	ATG Met	TTC Phe	CAT His 575	TAT		1728	
CTT Leu	CGT Arg	GCT Ala	CAA Gln 580	GAA Glu	TTT Phe	GAA Glu	CAT His	GGT Gly 585	AAA Lys	TCT Ser	AGG Arg	ATT Ile	GCT Ala 590	TTA Leu	ACA Thr		1776	
AAT Asn	TCT Ser	GTT Val 595	AAC Asn	GAA Glu	GCA Ala	TTA Leu	TTA Leu 600	Asn	CCT Pro	AGT Ser	CGT Arg	GTT Val 605	TAT Tyr	ACA Thr	TTT Phe		1824	
TTT Phe	TCT Ser 610	TCA Ser	GAC Asp	TAT Tyr	GTA Val	AAG Lys 615	AAA Lys	GTT Val	AAT Asn	AAA Lys	GCT Ala 620	ACG Thr	GAG Glu	GCA Ala	GCT Ala		1872	
625	Pne	TTA Leu	GIY	Trp	630	Glu	Gln	Leu	Val	Tyr 635	Asp	Phe	Thr	Asp	Glu 640		1920	•
Thr	Ser	GAA Glu	Val	Ser 645	Thr	Thr	Asp	Lys	Ile 650	Ala	Asp	Ile	Thr	Ile 655	Ile	:	1968	
ATT	CCA Pro	TAT	ATA Ile 660	Gly	Pro	GCT Ala	Leu	Asn	Ile	Gly	Asn	Met	Leu	Tyr	AAA Lys	:	2016	
Asp	GAT Asp	TTT Phe 675	GTA Val	GGT Gly	GCT Ala	Leu	ATA Ile 680	TTT Phe	TCA Ser	GGA Gly	GCT Ala	GTT Val 685	ATT Ile	CTG Leu	TTA Leu	:	2064	
GAA Glu	TTT Phe 690	ATA Ile	CCA Pro	GAG Glu	ATT Ile	GCA Ala 695	ATA Ile	CCT Pro	GTA Val	TTA Leu	GGT Gly 700	ACT Thr	TTT Phe	GCA Ala	CTT Leu	:	2112	
GTA Val 705	TCA Ser	TAT Tyr	ATT Ile	GCG Ala	AAT Asn 710	AAG Lys	GTT Val	CTA Leu	ACC Thr	GTT Val 715	CAA Gln	ACA Thr	ATA Ile	GAT Asp	AAT Asn 720	:	2160	
GCT Ala	TTA Leu	AGT Ser	AAA Lys	AGA Arg 725	AAT Asn	GAA Glu	AAA Lys	TGG Trp	GAT Asp 730	GAG Glu	GTC Val	TAT Tyr	AAA Lys	TAT Tyr 735	ATA Ile	2	2208	
GTA Val	ACA Thr	AAT Asn	TGG Trp 740	TTA Leu	GCA Ala	AAG Lys	Val	AAT Asn 745	ACA Thr	CAG Gln	ATT Ile	GAT Asp	CTA Leu 750	ATA Ile	AGA Arg	2	2256	

Lys	A.AAA E.Lys	ATC Met 755	. Lys	GAA Glu	GCT Ala	TTA Leu	GAA Glu 760	ASI	CAA Glr	A GCA n Ala	GAA Glu	A GCA 1 Ala 769	Thr	A AAG	G GCT s Ala		2304
ATA Ile	ATA Ile 770	. ASII	TAT	CAG Gln	TAT	AAT Asn 775	CAA Gln	TAT	ACT	GAG Glu	GAA Glu 780	i Glu	Lys	AA7 Asr	T AAT Asn		2352
ATT Ile 785	ASII	TTT Phe	AAT Asn	ATT	GAT Asp 790	GAT Asp	TTA Leu	AGT	TCG Ser	AAA Lys 795	CTT Leu	AAT Asn	GAG Glu	TCT	ATA Ile 800		2400 ⁻
ASII	гуs	Ala	мес	805	ASN	11e	ASN	Lys	Phe 810	Leu	Asn	Gln	Cys	Ser 815			2448
TCA Ser	TAT	TTA Leu	ATG Met 820	AAT Asn	TCT	ATG Met	ATC Ile	CCT Pro 825	TAT Tyr	GGT Gly	GTT Val	AAA Lys	CGG Arg 830	TTA Leu	GAA Glu		2496
GAT Asp	TTT Phe	GAT Asp 835	GCT Ala	AGT Ser	CTT Leu	AAA Lys	GAT Asp 840	GCA Ala	TTA Leu	TTA Leu	AAG Lys	TAT Tyr 845	ATA Ile	TAT Tyr	GAT Asp	:	2544
AAT Asn	AGA Arg 850	GGA Gly	ACT Thr	TTA Leu	ATT Ile	GGT Gly 855	CAA Gln	GTA Val	GAT Asp	AGA Arg	TTA Leu 860	AAA Lys	GAT Asp	AAA Lys	GTT Val	2	2592
AAT Asn 865	AAT Asn	ACA Thr	CTT Leu	Ser	ACA Thr 870	GAT Asp	ATA Ile	CCT Pro	Phe	CAG Gln 875	CTT Leu	TCC Ser	AAA Lys	TAC Tyr	GTA Val 880	2	2640
GAT A sp	AAT Asn	CAA Gln	AGA '	TTA Leu 885	TTA Leu	TCT . Ser '	ACA Thr	Phe	ACT Thr 890	GAA Glu	TAT Tyr	ATT Ile	Lys	TAA * 895		2	685

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 895 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu
1 5 10

Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn 20 25 30 -

Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro

Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys 50 55 60

Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly 65 70 75 80

Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr 85 90 95

Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys

Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg 120 Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val 155 Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile 165 Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly 185 His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr 250 Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala 260 Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr 280 Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met 330 Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu 360 Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu 375 Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn 395 Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg 405 Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn 425 Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe 440 Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser

Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val 470 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 490 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 505 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 520 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 555 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 565 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 585 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 600 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 630 635 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 650 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 680 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 695 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 730 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val

810

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 820 825 830

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 835 840 845

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 865 870 875 880

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys * 885 890 895

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2622 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2622

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGA Gly 1	TCC Ser	ATG Met	GAG Glu	TTC Phe 5	GTG Val	AAC Asn	AAG Lys	CAG Gln	TTC Phe 10	AAC Asn	TAT Tyr	AAG Lys	GAC Asp	CCT Pro 15	GTA Val	48
ASII	GGT Gly	vai	20	11e	Ala	Tyr	Ile	Lys 25	Ile	Pro	Lys	Tyr	Gly 30	Gln	Met	96
CAG Gln	CCG Pro	GTG Val 35	AAG Lys	GCT Ala	TTC Phe	AAG Lys	ATT Ile 40	CAT His	AAC Asn	AAA Lys	ATC Ile	TGG Trp 45	GTT Val	ATT Ile	CCG Pro	144
GAA Glu	CGC Arg 50	GAT Asp	ACA Thr	TTT Phe	ACG Thr	AAC Asn 55	CCG Pro	GAA Glu	GAA Glu	GGA Gly	GAC Asp 60	TTG Leu	AAC Asn	CCG Pro	CCG Pro	192
CCG Pro 65	GAA Glu	GCA Ala	AAG Lys	CAG Gln	GTG Val 70	CCA Pro	GTT Val	TCA Ser	TAC Tyr	TAC Tyr 75	GAT Asp	TCA Ser	ACC Thr	TAT Tyr	CTG Leu 80	240
AGC Ser	ACA Thr	GAC Asp	AAC Asn	GAG Glu 85	AAG Lys	GAT Asp	AAC Asn	TAC Tyr	CTG Leu 90	AAG Lys	GGA Gly	GTG Val	ACC Thr	AAA Lys 95	TTA Leu	288
TTC Phe	GAG Glu	CGT Arg	ATT Ile 100	TAT Tyr	TCC Ser	ACT Thr	GAC Asp	CTG Leu 105	GGC Gly	CGT Arg	ATG Met	CTG Leu	CTG Leu 110	ACC Thr	TCA Ser	336
ATC Ile	GTC Val	CGC Arg 115	GGA Gly	ATC Ile	CCA Pro	TTT Phe	TGG Trp 120	GGT Gly	GGC Gly	AGT Ser	ACC Thr	ATT Ile 125	GAC Asp	ACG Thr	GAG Glu	384
TTG Leu	AAG Lys 130	GTT Val	ATT Ile	GAC Asp	ACT Thr	AAC Asn 135	TGC Cys	ATT Ile	AAC Asn	GTG Val	ATC Ile 140	CAA Gln	CCA Pro	GAC Asp	GGT Gly	432

AG Se: 14!	r ry	C AC	SA TO	T GA r Gl	A GA u Gl 15	a re	T AA L As:	C CT n Le	C GT. u Va	A ATO	e II	C GG e Gl	G CC y Pr	C TC	CC GCG F Ala 160	•	480
GA(As _I	C AT	T AI e Il	C CA e Gl	G TT n Ph 16	e GI	G TGC u Cys	C AAG Ly:	G AG	C TT: r Phe 170	e Gly	CA Y Hi	C GA s Gl	A GT u Va	G TT l Le 17	G AAC u Asn 5		528
CT(Let	G AC	G CG r Ar	T AA g As 18	n GI	T TA	C GGC r Gly	TC: Sei	T ACT	r Glr	TAÇ 1 Tyı	Ile	T CG	T TTO g Pho 190	≥ Se	C CCA		576
Asp	PIN	19	5	e Gi	y Phe	e GIU	200	ı Ser	r Leu	ı Glu	ı Val	209	Th:	As	C CCG n Pro		624
CTG Leu	TT0 Let 210	r GT	T GC. y Al	A GGO a Gly	C AAC / Lys	TTC Phe 215	GCA Ala	ACT Thr	GAT Asp	CCA Pro	GCC Ala 220	a Val	ACC Thr	CTC	G GCA 1 Ala		672 ·
225	GI	ı re	1 116	e Hls	230	GIY	His	Arg	, Leu	Tyr 235	Gly	' Ile	Ala	Ile	AAC Asn 240		720
CCG Pro	AAC Asn	CGC Arg	GTC Val	TTC Phe 245	. Lys	GTT Val	AAC Asn	ACC Thr	AAC Asn 250	GCC Ala	TAC	TAC	GAG Glu	ATO Met 255			768
GGT Gly	TTA Leu	GAA Glu	GTA Val 260	. Ser	TTC Phe	GAG Glu	GAA Glu	CTG Leu 265	CGC Arg	ACG	TTC Phe	GGT Gly	GGC Gly 270	CAT His	GAT Asp		816
GCG Ala	AAG Lys	TTT Phe 275	TIE	GAC Asp	AGC Ser	TTG Leu	CAG Gln 280	GAG Glu	AAC Asn	GAG Glu	TTC Phe	CGT Arg 285	CTG Leu	TAC Tyr	TAC Tyr		864
TAC Tyr	AAC Asn 290	AAG Lys	TTT Phe	AAA Lys	GAT Asp	ATT Ile 295	GCA Ala	AGT Ser	ACA Thr	CTG Leu	AAC Asn 300	AAG Lys	GCT Ala	AAG Lys	TCC Ser		912
ATT Ile 305	GTG Val	GGT Gly	ACC Thr	ACT Thr	GCT Ala 310	TCA Ser	TTA Leu	CAG Gln	TAT Tyr	ATG Met 315	AAA Lys	AAT Asn	GTT Val	TTT Phe	AAA Lys 320		960
GAG Glu	AAA Lys	TAT Tyr	CTC Leu	CTA Leu 325	TCT Ser	GAA Glu	Asp	Thr	Ser	GGA Gly	Lys	Phe	Ser	GTA Val 335	Asp		1008
AAA Lys	TTA Leu	AAA Lys	TTT Phe 340	GAT Asp	AAG Lys	TTA Leu	TAC Tyr	AAA Lys 345	ATG Met	TTA Leu	ACA Thr	GAG Glu	ATT Ile 350	TAC Tyr	ACA Thr	;	1056
GAG Glu	GAT Asp	AAT Asn 355	TTT Phe	GTT Val	AAG Lys	TTT Phe	TTT Phe 360	AAA Lys	GTA Val	CTT . Leu .	AAC Asn	AGA Arg 365	AAA Lys	ACÁ Thr	TAT Tyr	•	1104
Leu .	AAT Asn 370	TTT Phe	GAT Asp	AAA Lys	GCC Ala	GTA ' Val 1 375	TTT Phe	AAG Lys	ATA . Ile .	Asn :	ATA Ile 380	GTA Val	CCT Pro	AAG Lys	GTA Val	:	1152
AAT Asn 385	TAC Tyr	ACA Thr	ATA Ile	TAT Tyr	GAT Asp 390	GGA :	Phe	AAT Asn	Leu i	AGA 1 Arg 1 395	AAT Asn	ACA .	AAT : Asn :	TTA Leu	GCA Ala 400	1	1200
GCA A	AAC Asn	TTT Phe	AAT Asn	GGT Gly 405	CAA Gln	AAT A Asn 1	ACA (Thr (Glu	ATT I Ile I 410	AAT <i>l</i> Asn <i>l</i>	AAT . Asn i	ATG A	Asn i	TTT Phe	ACT Thr	1	.248

AAA Lys	CTA Leu	AAA Lys	AAT Asn 420	TTT Phe	ACT Thr	GGA Gly	TTG Leu	TTT Phe 425	GAA Glu	TTT	TAT Tyr	AAG Lys	TTG Leu 430	CTA Leu	TGT Cys	1296
GTA Val	AGA Arg	GGG Gly 435	ATA Ile	ATA Ile	ACT Thr	TCT. Ser	AAA Lys 440	ACT Thr	AAA Lys	TCA Ser	TTA Leu	GAT Asp 445	Lys	GGA Gly	TAC	1344
AAT Asn	AAG Lys 450	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 455	TGT Cys	ATC Ile	AAA Lys	GTT Val	AAT Asn 460	AAT Asn	TGG Trp	GAC Asp	TTG Leu	1392
TTT Phe 465	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 470	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 475	GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 480	1440
GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 485	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 490	GCA Ala	GCA Ala	GAA Glu	GAA Glu	AAT Asn 495	ATT Ile	1488
					CAA Gln											1536
GAA Glu	CCT Pro	GAA Glu 515	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 520	AAT Asn	CTT Leu	TCA Ser	AGT Ser	GAC Asp 525	ATT Ile	ATA Ile	GGC Gly	1584
					CCT Pro											1632
					TAT Tyr 550											1680
					TCT Ser											1728
					AGT Ser											1776
GTA Val	AAG Lys	AAA Lys 595	GTT Val	AAT Asn	AAA Lys	GCT Ala	ACG Thr 600	GAG Glu	GCA Ala	GCT Ala	ATG Met	TTT Phe 605	TTA Leu	GGC Gly	TGG Trp	1824
GTA Val	GAA Glu 610	CAA Gln	TTA Leu	GTA Val	TAT Tyr	GAT Asp 615	TTT Phe	ACC Thr	GAT Asp	GAA Glu	ACT Thr 620	AGC Ser	GAA Glu	GTA Val	AGT Ser	1872
ACT Thr 625	ACG Thr	GAT Asp	AAA Lys	ATT Ile	GCG Ala 630	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 635	ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 640	1920
CCT Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 645	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 650	AAA Lys	GAT Asp	GAT Asp	TTT Phe	GTA Val 655	GGT Gly	1968
GCT Ala	TTA Leu	ATA Ile	TTT Phe 660	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 665	CTG Leu	TTA Leu	GAA Glu	TTT Phe	ATA Ile 670	CCA Pro	GAG Glu	2016
ATT Ile	GCA Ala	ATA Ile 675	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 680	TTT Phe	GCA Ala	CTT Leu	GTA Val	TCA Ser 685	TAT Tyr	ATT Ile	GCG Ala	2064

AA? Ası	1 Ly:	o va	T CT. l Le	A AC u Th	C GT r Va	r CAA l Glr 699	, LUI	A ATA	A GAT	AA:	r GC: n Ala 700	a Lei	A AG	T AA.	A AGA s Arg	2112
AAT Asr 705	1 61	A AA.	A TGO	G GA' P As	F GAG P Glu 710	ı vaı	TA1	Lys	TAT	T ATA T Ile 715	e Val	A ACA	AAT Asi	r TG(G TTA Leu 720	2160
MIG	гуs	va.	L ASI	725	GI.	ıııe	Asp	Leu	730	Arc	Lys	Lys	Met	735		2208
Ala	ren	GIL	740)	ı Ala	Glu	Ala	745	Lys	Ala	Ile	Ile	Asn 750	Tyr	CAG Gln	2256
Lyr	ASI	755	Tyr	ŢŊĬ	Giu	Glu	760	Lys	. Asn	Asn	Ile	As n 765	Phe	Asn	ATT Ile	2304
Asp	770	ren	ser	Ser	цуs	775	Asn	GIu	Ser	Ile	Asn 780	AAA Lys	Ala	Met	Ile	2352
785	iie	ASN	ьуѕ	Pne	790	Asn	GIn	Cys	Ser	Val 795	Ser	TAT Tyr	Leu	Met	Asn 800	2400
ser	Mec	iie	Pro	805	GIĀ	Val	Lys	Arg	Leu 810	Glu	Asp	TTT Phe	Asp	Ala 815	Ser	2448
CTT Leu	AAA Lys	GAT Asp	GCA Ala 820	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 825	TAT Tyr	GAT Asp	AAT Asn	AGA Arg	GGA Gly 830	ACT Thr	TTA Leu	2496
ATT Ile	GGT Gly	CAA Gln 835	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 840	GAT Asp	AAA Lys	GTT Val	AAT Asn	AAT Asn 845	ACA Thr	CTT Leu	AGT Ser	25 4 4
inr	GAT Asp 850	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 855	TCC . Ser	AAA Lys	TAC Tyr	Val	GAT Asp 860	AAT Asn	CAA Gln	AGA Arg	TTA Leu	2592
rta Leu 165	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 870	TAT . Tyr	ATT . Ile :	AAG Lys	TAA *							2622

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 874 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val

Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met 30

Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro

Glu	Arg 50	Asp	Thr	Phe	Thr	Asn 55	Pro	Glu	Glu	Gly	Asp 60	Leu	Asn	Pro	Pro
Pro 65	Glu	Ala	Lys	Gln	Val 70	Pro	Val	Ser	Tyr	Tyr 75	Asp	Ser	Thr	Tyr	Leu 80
Ser	Thr	Asp	Asn	Glu 85	Lys	Asp	Asn	Tyr	Leu 90	Lys	Gly	Val	Thr	Lys 95	Leu
Phe	Glu	Arg	Ile 100	Tyr	Ser	Thr	Asp	Leu 105	Gly	Arg	Met	Leu	Leu 110	Thr	Ser
Ile	Val	Arg 115	Gly	Ile	Pro	Phe	Trp 120	Gly	Gly	Ser	Thr	Ile 125	Asp	Thr	Glu
Leu	Lys 130	Val	Ile	Asp	Thr	Asn 135	Cys	Ile	Asn	Val	Ile 140	Gln	Pro	Asp	Gly
Ser 145	Tyr	Arg	Ser	Glu	Glu 150	Leu	Asn	Leu	Val	Ile 155	Ile	Gly	Pro	Ser	Ala 160
Asp	Ile	Ile	Gln	Phe 165	Glu	Cys	Lys	Ser	Phe 170	Gly	His	Glu	Val	Leu 175	Asn
Leu	Thr	Arg	Asn 180		Tyr	Gly		Thr 185	Gln	Tyr	Ile	Arg	Phe 190		Pro
Asp	Phe	Thr 195	Phe	Gly	Phe	Glu	Glu 200	Ser	Leu	Glu	Val	Asp 205	Thr	Asn	Pro
Leu	Leu 210	Gly	Ala	Gly	Lys	Phe 215	Ala	Thr	Asp	Pro	Ala 220	Val	Thr	Leu	Ala
His 225	Glu	Leu	Ile	His	Ala 230	Gly.	His	Arg	Leu	Tyr 235	Gly	Ile	Ala	Ile	Asn 240
Pro	Asn	Arg	Val	Phe 245	Lys	Val	Asn	Thr	Asn 250	Ala	Tyr	Tyr	Glu	Met 255	Ser
Gly	Leu	Glu	Val 260	Ser	Phe	Glu	Glu	Leu 265	Arg	Thr	Phe	Gly	Gly 270	His	Asp
Ala	Lys	Phe 275	Ile	Asp	Ser	Leu	Gln 280	Glu	Asn	Glu	Phe	Arg 285	Leu	Tyr	Tyr
Tyr	Asn 290	_	Phe	Lys	Asp	Ile 295	Ala	Ser	Thr	Leu	Asn 300	Lys	Ala	Lys	Ser
11e 305	Val	Gly	Thr	Thr	Ala 310	Ser	Leu	Gln	Tyr	Met 315	Ŀуs	Asn	Val	Phe	Lys 320
Glu	Lys	Tyr	Leu	Leu 325	Ser	Glu	Asp	Thr	Ser 330	Gly	Lys	-Phe	Ser	Val 335	Asp
Lys	Leu	Lys	Phe 340	Asp	Lys	Leu	Tyr	Lys 345	Met	Leu	Thr	Glu	11e 350	Tyr	Thr
Glu	Asp	Asn 355	Phe	Val	Lys	Phe	Phe 360	Lys	Val	Leu	Asn	Arg 365	Lys	Thr	Tyr
Leu	Asn 370	Phe	Asp	Lys	Ala	Val 375	Phe	Lys	Ile	Asn	Ile 380	Val	Pro	Lys	Val
Asn 385	Tyr	Thr	Ile	Tyr	Asp	Gly	Phe	Asn	Leu	Arg 395	Asn	Thr	Asn	Leu	Ala 400

Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys 425 Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly 475 Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile 490 Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys 535 Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu 570 Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly 630 Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu 665 Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala 680 Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln

745

Tyr	Asn	Gln 755	Tyr	Thr	Glu	Glu	Glu 760	Lys	Asn	Asn	Ile	Asn 765	Phe	Asn	Ile
Asp	Asp 770	Leu	Ser	Ser	Lys	Leu 775	Asn	Glu	Ser	Ile	Asn 780	Lys	Ala	Met	Ile
Asn 785	Ile	Asn	Lys	Phe	Leu 790	Asn	Gln	Cys	Ser	Val 795		Tyr	Leu	Met	Asn 800
Ser	Met	Ile	Pro	Tyr 805	Gly	Val	Lys	Arg	Leu 810	Glu	Asp	Phe	Asp	Ala 815	Ser
Leu	Lys	Asp.	Ala 820	Leu	Leu	Lys	Tyr	Ile 825		Asp	Asn	Arg	Gly 830	Thr	Leu
Ile	Gly	Gln 835	Val	Asp	Arg	Leu	Lys 840	Asp	Lys	Val	Asn	Asn 845	Thr	Leu	Ser
Thr	Asp 850	Ile	Pro	Phe	Gln	Leu 855	Ser	Lys	Tyr-	Val	Asp 860	Asn	Gln	Arg	Leu
Leu 865	Ser	Thr	Phe	Thr	Glu 870	Tyr	Ile	Lys	*	•					
(2)	INFO	ORMAT	CION	FOR	SEQ	ID 1	10:	7: .							
	(i)	SE	QUEN	CE CI	LARAC	CTER	ISTIC	: 2S		*					

- (A) LENGTH: 2613 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:

 - (A) NAME/KEY: CDS
 (B) LOCATION:1..2613
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATG Met 1	CCA Pro	TTT Phe	GTT Val	AAT Asn 5	AAA Lys	CAA Gln	TTT Phe	AAT Asn.	TAT Tyr 10	AAA Lys	GAT Asp	CCT Pro	GTA Val	AAT Asn 15	GGT Gly	4	В
GTT Val	GAT Asp	ATT Ile	GCT Ala 20	TAT Tyr	ATA Ile	AAA Lys	ATT Ile	CCA Pro 25	AAT Asn	GCA Ala	GGA Gly	CAA Gln	ATG Met 30	CAA Gln	CCA Pro	9	6
GTA Val	AAA Lys	GCT Ala 35	TTT Phe	AAA Lys	ATT Ile	CAT His	AAT Asn 40	AAA Lys	ATA Ile	TGG Trp	GTT Val	ATT Ile 45	CCA Pro	GAA Glu	AGA Arg	14	4
GAT Asp	ACA Thr 50	TTT Phe	ACA Thr	AAT Asn	CCT Pro	GAA Glu 55	GAA Glu	GGA Gly	GAT Asp	TTA Leu	AAT Asn 60	CCA Pro	CCA Pro	CCA Pro	GAA Glu	19	2
GCA Ala 65	AAA Lys	CAA Gln	GTT Val	CCA Pro	GTT Val 70	TCA Ser	TAT Tyr	TAT Tyr	GAT Asp	TCA Ser 75	ACA Thr	TAT Tyr	TTA Leu	AGT Ser	ACA Thr 80	24	0
GAT Asp	AAT Asn	GAA Glu	AAA Lys	GAT Asp 85	AAT Asn	TAT Tyr	TTA Leu	AAG Lys	GGA Gly 90	GTT Val	ACA Thr	AAA Lys	TTA Leu	TTT Phe 95	GAG Glu	28	8

AG/ Arg	A ATT	TATE Tyr	TCA Ser 100	Thi	GAT Asp	CTI Leu	GGA Gly	AGA Arg 105	, Met	TTC Leu	TT)	A AC	A TC	r Il	A GTA e Val	336
AG(G GG# G Gly	ATA Ile 115	Pro	TTT Phe	TGG	G GGT	GGA Gly 120	Ser	ACA Thr	ATA Ile	GAT Asp	T ACA	Gl	A TT.	A AAA u Lys	384
GT1 Val	Ile 130	Asp	ACT Thr	TAA '	TGI Cys	ATT Ile 135	AAT Asn	GTG Val	Ile	CAA Gln	CCA Pro 140	Asp	GGT Gly	r AGʻ ⁄ Se:	r TAT	432
AGA Arg 145	Ser	GAA Glu	GAA Glu	CTT Leu	AAT Asn 150	Leu	GTA Val	ATA	ATA Ile	GGA Gly 155	Pro	TCA Ser	GCT Ala	GA:	T ATT D Ile 160	480
ATA Ile	CAG Gln	TTT Phe	GAA Glu	TGT Cys 165	Lys	AGC Ser	TTT Phe	GGA Gly	CAT His 170	Glu	GTT Val	TTG Leu	AA1 Asn	CT1 Let 175	ACG Thr	528
CGA Arg	AAT Asn	GGT Gly	TAT Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAA Gln	TAC Tyr 185	ATT Ile	AGA Arg	TTT Phe	AGC Ser	CCA Pro 190	Asp	TTT Phe	576
ACA Thr	TTT Phe	GGT Gly 195	TTT Phe	GAG Glu	GAG Glu	TCA Ser	CTT Leu 200	GAA Glu	GTT Val	GAT Asp	ACA Thr	AAT Asn 205	CCT Pro	CTT Leu	TTA Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAA Lys	TTT Phe	GCT Ala	ACA Thr 215	GAT Asp	CCA Pro	GCA Ala	GTA Val	ACA Thr 220	TTA Leu	GCA Ala	CAT His	GAA Glu	672
CTT Leu 225	ATA Ile	CAT His	GCT Ala	GGA Gly	CAT His 230	AGA Arg	TTA Leu	TAT Tyr	GGA Gly	ATA Ile 235	GCA Ala	ATT Ile	AAT Asn	CCA Pro	AAT Asn 240	720
AGG Arg	GTT Val	TTT Phe	AAA Lys	GTA Val 245	TAA neA	ACT Thr	AAT Asn	GCC Ala	TAT Tyr 250	TAT Tyr	GAA Glu	ATG Met	AGT Ser	GGG Gly 255	TTA Leu	768 [°]
Glu	Val	Ser	Phe 260	Glu	GAA Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys	81 6
TTT Phe	ATA Ile	GAT Asp 275	AGT Ser	TTA Leu	CAG Gln	GAA Glu	AAC Asn 280	GAA Glu	TTT Phe	CGT Arg	CTA Leu	TAT Tyr 285	TAT Tyr	TAT Tyr	AAT Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATA Ile	GCA Ala	AGT Ser 295	ACA Thr	CTT Leu	AAT Asn	AAA Lys	GCT Ala 300	AAA Lys	TCA Ser	ATA Ile	GTA Val	912
GGT Gly 305	ACT Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	Ser	GAA Glu 325	GAT A sp	ACA Thr	TCT (Ser	Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	Met :	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT A sp	1056
					TTT Phe	Lys										1104

TTT Phe	GAT Asp 370	rys	GCC Ala	GTA Val	TTI Phe	AAG Lys 375	TTe	AAT Asn	ATA	GTA Val	CCT Pro	Lys	GŢA Val	AA1 Asr	TAC Tyr		1152
385	116	lyr	Asp	Gly	390	Asn	Leu	Arg	Asn	395	Asn	Let	Ala	Ala	AAC Asn 400	:	1200
Pne	Asn	GIY	GIN	405	THE	GIU	lle	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415		:	1248
rys	Asn	Pne	420	Gly	ren	Pne	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	AGA Arg	3	1296
GIY	Ile	11e 435	Thr	Ser	гÀг	ACT Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys	1	
Ala	Leu 4 50	Asn	Asp	Leu	Cys	ATC Ile 455	Lys	Val	Asn	Asn	Trp 460	Asp	Leu	Phe	Phe	1	1392
Ser 465	Pro	Ser	Glu	Asp	Asn 470	TTT Phe	Thr	Asn	Asp	Leu 475	Asn	Lys	Gly	Glu	Glu 480	1	440
Ile	Thr	Ser	Asp	Thr 485	Asn	ATA Ile	Glu	Ala	Ala 490	Glu	Glu	Asn	Ile	Ser 495	Leu	1	488
Asp	Leu	Ile	Gln 500	Gln	Tyr	TAT Tyr	Leu	Thr 505	Phe	Asn	Phe	Asp	Asn 510	Glu	Pro	1	.536
Glu	Asn.	Ile 515	Ser	Ile	Glu	AAT Asn	Leu 520	Ser	Ser	Asp	Ile	Ile 525	Gly	Gln	Leu	1	584
Glu	Leu 530	Met	Pro	Asn	Ile	GAA Glu 535	Arg	Phe	Pro	Asn	Gly 540	Lys	Lys	Tyr	Glu	1	632
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC	CAT	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	1	680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 525	TTA Leu	1	728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys	1	776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu	1	824
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	11	872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	19	920

TT <i>I</i> Let	A AAT	r ATA	A GGT e Gl	AA: AS: 64:	n met	G TTA	A TA	T AAI r Ly:	A GA S As 65	P As	T TT p Ph	T GT e Va	A GO 1 GI	ST GO Ly Al	TTA La Leu	1	1968
ATA Ile	TTT Phe	TC/ Sei	A GGA C Gly 660	Ala	GT7 Val	T ATT	CTO Let	G TTA u Let 669	ı Glu	A TT	r at	A CC	A GA o Gl 67	u []	T GCA e Ala	. .	2016
ATA Ile	CCT Pro	GTA Val 675	. Leu	GG1 Gly	ACT Thr	TTT Phe	GC2 Ala 680	a Leu	GT/ Val	A TCA	A TA	T AT: 116	e Al	G AA a As	T AAG n Lys		2064
GTT Val	CTA Leu 690	Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AG: Sei 700	c Lys	A AG	A AA g As	T GAA n Glu		2112
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	Lys	TAT Tyr	T ATA	GTA Val	ACA Thr 715	Asr	TGO	TT:	A GC. u Al	A AAG a Lys 720		2160
GTT Val	AAT Asn	ACA Thr	CAG Gln	ATT Ile 725	Asp	CTA Leu	ATA	AGA Arg	AAA Lys 730	Lys	ATC Met	AAA Lys	GA/	A GC: 1 Ala 735	TTA Leu		2208
GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT	CAC Glr 750	Ty	AAT Asn		2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATI	GAT Asp	GAT Asp		2304
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA		2352
AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800		2400
ATC Ile	CCT Pro	TAT Tyr	Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys		2448
GAT (GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT . Tyr	ATA '	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly		2496
CAA (Gln)	Val .	GAT Asp 835	AGA '	TTA . Leu :	AAA (Lys)	Asp 1	AAA Lys 840	GTT . Val .	AAT . Asn .	AAT . Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp		2544
ATA (CCT : Pro :	TTT Phe	CAG (Gln)	CTT :	Ser 1	AAA : Lys :	rac Fyr	GTA (Val	GAT Asp	Asn (CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser		2592
ACA 1 Thr E 865				lyr :							i						2613

(2) INFORMATION FOR SEQ ID NO: 8:

⁽i) SEQUENCE CHARACTERISTICS:

⁽A) LENGTH: 871 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro .25 Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 105 Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 150 Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 200 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 250 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 315

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435
440
445

Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 460

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 465 470 475 480

Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495

Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 505 510

Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 590

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655

Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685

Val	Leu 690	Thr	Val	Gln	Thr	Ile 695	Asp	Asn	Ala	Leu	Ser 700	Lys	Arg	Asn	Glu		
Lys 705	Trp	Asp	Glu	Val	Tyr 710	Lys	Tyr	Ile	Val	Thr 715	Asn	Trp	Leu	Ala	Lys 720		
Val	Asn	Thr	Gln	Ile 725	Asp	Leu	Ile	Arg	Lys 730	Lys	Met	Lys	Glu	Ala 735	Leu		
Glu	Asn	Gln	Ala 740	Glu	Ala	Thr	Lys	Ala 745	Ile	Ile	Asn	Tyr	Gln 750	Tyr	Asn		
Gln	Tyr	Thr 755	Glu	Glu	Glu	Lys	Asn 760	Asn	Ile	Asn	Phe	Asn 765	Ile	Asp	Asp		
Leu	Ser 770	Ser	Lys	Leu	Asn	Glu 775	Ser	Ile	Asn	Lys	Ala 780	Met	Ile	Asn	Ile		
Asn 785	Lys	Phe	Leu	Asn	Gln 790	Cys	Ser	Val	Ser	Tyr 795	Leu	Met	Asn	Ser	Met 800		
Ile	Pro	Tyr	Gly	Val 805	Lys	Arg	Leu	Glu	Asp 810	Phe	Asp	Ala	Ser	Leu 815	Lys		
Asp	Ala	Leu	Leu 820	Lys	Tyr	Ile	Tyr	Asp 825	Asn	Arg	Gly	Thr	Leu 830	Ile	Gly		
Gln	Val	Asp 835	Arg	Leu	Lys	Asp	Lys 840	Val	Asn	Asn	Thr	Leu 845	Ser	Thr	Asp		
Ile	Pro 850	Phe	Gln	Leu	Ser	Lys 855	Tyr	Val	Asp	Asn	Gln 860	Arg	Leu	Leu	Ser		
Thr 865	Phe	Thr	Glu	Tyr	Ile 870	Lys											
(2)	INF	ORMA'	rion	FOR	SEQ	ID I	10:	9:									
	(i)	() () ()	A) LI B) Ti C) Si	CE CI ENGTI YPE: IRANI OPOLO	nuc DEDN	628 leic ESS:	base aci dou	pai: d	rs								
				LE T	YPE:	DNA	(ge	nomi	c)								
	(ix	. (.	ATURI A) Ni B) L	E: AME/ OCAT	KEY : ION :	CDS	628		•								
	(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0: 9	:						
ATG Met	Gln	TTC Phe	GTG Val	AAC Asn 5	AA G Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly		48
GTI Val	GAC Asp	ATT	GCC Ala 20	Tyr	ATC Ile	AAA Lys	ATT Ile	CCA Pro 25	Asn	GCC	GGC Gly	CAG Gln	ATG Met 30	Gln	CCG Pro		96
GTC Val	AAG Lys	GCT Ala 35	Phe	AAG Lys	ATT Ile	CAT His	AAC Asr	ı Lys	ATC	TGG	GTT Val	ATT Ile 45	Pro	GAA Glu	CGC Arg	1	44
GAT Ası	T ACA	Phe	ACG Thr	AAC Asr	CCC Pro	GAA Glu 55	ı Glı	A GGA 1 Gly	GAC Asp	TTG Lev	AAC Asr 60	Pro	CCG Pro	CCG	GAA Glu	1	9 2

GCA Ala 69	a Lys	G CA	G GTO	G CC.	A GTT O Val	L Sei	TAC Ty	TAC TY	C GA	TC Se: 7	r Th	C TA	T CI	G AG	C ACA r Thr	240
GA(Asp	AAC Asr	GA(G AAG	G GA S Asi 89) Asr	TAC Tyr	CTC Lev	AA(Ly:	G GG/ s Gl ₃ 90	/ Val	G AC	C AA r Ly	A.TT s Le	u Ph	C GAG e Glu 5	288
CGT Arg	ATT	TA:	TCC Ser 100	Thi	GAC Asp	CTG Leu	GGC Gly	CG7 Arg 105	g Met	CTC Lev	CTO	G AC u Th	C TC r Se ll	r Il	C GTC e Val	336
CGC Arg	GGA Gly	ATC 11e	Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	Sei	T ACC	ATT	GAG Asi	2 AC	r Gl	G TT	G AAG u Lys	384
GTT Val	Ile 130	Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro	Asp	G GG G Gly	r AG	C TAC	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC	ATC Ile	GGG Gly 155	Pro	TCO Ser	GCC Ala	G GA(C ATT D Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	Leu	AAC Asr	CTC Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	Pro	CTG Leu	TTG	624 .
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG (Glu .	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT	GCA Ala	AGT . Ser ' 295	ACA Thr	CTG Leu	AAC A sn	Lys .	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT . Gly ' 305	ACC Thr	ACT Thr	GCT Ala	Ser	TTA (Leu (310	CAG ' Gln '	TAT . Tyr !	ATG Met	Lys .	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT (CTC Leu	CTA Leu	Ser	GAA Glu 325	GAT A	ACA '	TCT (Ser (Gly	AAA Lys 330	TTT '	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008

AA/ Lys	TTI Phe	GAT Asp	AAC Lys	שבי	A TAC	Lys	ATC Met	TT/ Let 349	ı Tnı	GAC Glu	G ATT	TAC Tyr	C AC	r Gl	G GAT	1056
AAT Asn	TTT Phe	GTI Val 355	. Lys	Phe	TTT Phe	AAA Lys	GTA Val 360	rer	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	Ty	TTC Lev	G AAT 1 Asn	1104
TTT Phe	GAT Asp 370	Lys	GCC Ala	GTA Val	TTT Phe	Lys 375	Ile	AA1 Asn	ATA	GTA Val	CCT Pro 380	Lys	GTA Val	AA1 Asi	TAC Tyr	1152
ACA Thr 385	TIE	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr	Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
AGC Ser	GCT Ala 450	GAT Asp	GGG Gly	GCA Ala	TTA Leu	AAT Asn 455	GAT Asp	TTA Leu	TGT Cys	ATC Ile	AAA Lys 460	GTT Val	AAT Asn	AAT Asn	TGG Trp	1392
GAC Asp 465	ŤTG Leu	TTT Phe	TTT Phe	AGT Ser	CCT Pro 470	TCA Ser	GAA Glu	GAT Asp	AAT Asn	TTT Phe 475	ACT Thr	AAT Asn	GAT Asp	CTA Leu	AAT Asn 480	1440
AAA Lys	GGA Gly	GAA Glu	Glu	ATT Ile 485	ACA Thr	TCT Ser	GAT Asp	ACT Thr	AAT Asn 490	ATA Ile	GAA Glu	GCA Ala	GCA Ala	GAA Glu 495	GAA Glu	1488
AAT Asn	ATT Ile	AGT Ser	TTA Leu 500	GAT Asp	TTA Leu	ATA Ile	CAA Gln	CAA Gln 505	TAT Tyr	TAT Tyr	TTA Leu	ACC Thr	TTT Phe 510	AAT Asn	TTT Phe	1536
GAT Asp	AAT Asn	GAA Glu 515	CCT Pro	GAA Glu	AAT Asn	ATT Ile	TCA Ser 520	ATA Ile	Glu	AAT Asn	Leu	TCA Ser 525	AGT Ser	GAC Asp	ATT Ile	1584
ATA Ile	GGC Gly 530	CAA Gln	TTA Leu	GAA Glu	CTT Leu	ATG Met 535	CCT Pro	AAT Asn	ATA Ile	GAA Glu	AGA Arg 540	TTT Phe	CCT Pro	AAT Asn	GGA Gly	1632
AAA Lys 545	AAG Lys	TAT Tyr	GA G Glu	TTA Leu	GAT Asp 550	AAA Lys	TAT Tyr	ACT Thr	Met	TTC Phe 555	CAT His	TAT Tyr	CTT Leu	CGT Arg	GCT Ala 560	1680
CAA Gln	GAA Glu	TTT Phe	GAA Glu	CAT His 565	GGT Gly	AAA Lys	TCT Ser	AGG Arg	ATT Ile 570	GCT Ala	TTA Leu	ACA Thr	AAT Asn	TCT Ser 575	GTT Val	1728
AAC Asn	GAA Glu	GCA Ala	TTA Leu 580	TTA Leu	AAT Asn	CCT Pro	Ser	CGT Arg 585	GTT Val	TAT Tyr	ACA Thr	TTT Phe	TTT Phe 590	TCT Ser	TCA Ser	1776
GAC Asp	TAT Tyr	GTA Val 595	AAG Lys	AAA Lys	GTT Val	AAT Asn	AAA Lys 600	GCT Ala	ACG Thr	GAG Glu	Ala	GCT Ala 605	ATG Met	TTT Phe	TTA Leu	1824

	GG G1	, -	GG(rp) 10	GTA Val	GA Gl	A Ci u Gi	AA T		TA Tal T	TAT (Tyr)	GAT Asp	TT7 Phe	T AC ⇒ Th	r As	AT (sp (GAA Glu	ACT Thi	r AG	iC :r	GAA Glu		1872
	62	5		••••			AT AN Sp Ly 63	30	IC H	ııa F	sp	TIE	63:	r II 5	le 1	lle	Ile	Pr	0	Tyr 640		1920
			. y .	10	AIC	64	_	1.	re G	ту н	sn .	мет 650	Let	т Ту	r L	ys	Asp	As ₁	p : 5	Phe		1968
			, ,		660		A TT e Ph		-1 G	1 y A	65	vaı	ile	: Le	u L	eu	Glu 670	Phe	2]	lle	:	2016
			6	75	nia		A CC e Pr	O Va	68	30	TY.	inr	Phe	A1.	a L	eu 85	Val	Ser	7	yr	i	2064
		69	0	J.,	Бүз	va.	r CT.	69	5	II G.	ın ı	nr	He	700	o As O	sn 1	Ala	Leu	S	er	. 2	112
•	705	***	g As) ii	31u	Буз	710 710) AS	b GI	u Va	AI T	yr	Lys 715	Туг	: Il	le i	/al	Thr	7	sn 20	2	160
1	rp (GG	TT	A GC	CA I	AAG Lys	GTI Val 725	AA7 Asi	T AC.	A CA	G A1	e A	AT sp 30	CTA Leu	ATA Ile	A AG	A A	ys	AAA Lys 735	A' M	TG et	2	208
•	, y S	GIC	. 77	7	740	GIU	AAT Asn	GII	1 AL	74	u A. 5	la '	Thr	Lys	Al	a I 7	le 50	Ile	As	sn	2:	256
T	AT yr	CAG Gln	TA Ty 75		AT sn	CAA Gln	TAT Tyr	ACT Thi	GA(Gl) 76(T GT	A G	AG A	AAA Lys	AAT Asn	AA As: 76	n I	TT I	AAT Asn	TT	TT ne	2:	304
A	AT sn	ATT Ile 770	ns,	T G p A	SAT	TTA Leu	AGT Ser	TCG Ser 775	rys	CT:	I AA	AT C	Slu	TCT Ser 780	ATA Ile	A A	AT A	AAA .ys	GC Al	T . a	23	152
7.51	rg et 85	ATT Ile	AA' Ası	T A	TA . le .	AAT Asn	AAA Lys 790	TTT	TTC	AA1 Asi	CA 1 Gl	n C	GC Ys '95	TCT Ser	GTT Val	T TO	CA I	уr	TT Le 80	u	24	00
A: Me	rg .	AAT Asn	TC1	C A	ec.	ATC Ile 805	CCT Pro	TAT Tyr	GGT Gly	GTT Val	AA Ly 81	s A	GG :	ITA Leu	GAA Glu	A GA	p P	TT (he /	GA: As:	T p	24	48
G(A)	T .	AGT Ser	CTT		AA (ys <i>)</i> 20	TAE Asp	GCA Ala	TTA Leu	TTA Leu	AAG Lys 825	Ту	T A r I	TA 1	rat Tyr	GAT Asp	As	T A n A 0	GA (rg (GG: Gl:	A Y	24	96
AC Th	T 1	rTA Leu	ATT Ile 835	. 61	T C	CAA Sln	GTA Vạl	GAT Asp	AGA Arg 840	TTA Leu	AA. Ly:	A G	AT A sp I	yys	GTT Val 845	AA As	T A	AT A	AC#	A :	25	44
CT Le	u	AGT Ser 350	ACA Thr	GA As	T A	le [le	CCT Pro	TTT Phe 855	CAG Gln	CTT Leu	TC(Se	C Al	ys T	AC Yr 60	GTA Val	GA [*]	T AZ P As	AT C	AA Slr	A 1	259	92
AG Ar 86	g I	TTA Leu	TTA Leu	TC Se	T A	hr	TTT Phe 870	ACT Thr	GAA Glu	TAT Tyr	ATT Ile	A. A. A. B. T. B.	/S	AA *		-					262	8

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 876 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys 330

Phe Ser Val Asp Lys Leu 335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 370 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
420 425 430

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445

Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp 450 455 460

Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn 465 470 475 480

Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu 485 490 495

Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe 500 505 510

Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile 515 520 525

Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly
530 535 540

Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala 545 550 555 560

Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val 565 570 575

Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe-Phe Ser Ser 580 585 590

Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu 595 600 605

Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu 610 620

Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr 625 630 635 640

Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe 645 650 655

Val	Gly	Ala	Leu 660	Ile	Phe	Ser	Gly	Ala 665	Val	Ile	Leu	Leu	Glu 670	Phe	Ile
Pro	Glu	Ile	Ala	Ile	Pro	Val	Leu	Gly	Thr	Phe	Ala	Leu	Val	Ser	Tvr

675 680 The Ala Leu Val Ser Tyr

Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser 690 695 700

Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn 715 710 715

Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met
725 730 735

Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn 740 745 750

Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe 755 760 765

Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala 770 775 780

Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu 785 790 795 800

Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp 805 810 815

Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly 820 825 830

Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr 835 840 845

Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln 850 855 860

Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 865 870 875

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2637 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear
- (ii) MOLECULE_TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2637

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT
Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

96

PCT/GB97/02273

GTC Val	G AAG	2 71	T TI la Ph	CC AF	AG AT	TT CA le Hi	5 A:	AC A sn L	AA A ys I	TC le	TGG Trp	GT:	1 11	T C e P	CG G	AA lu	CGC Arg		144
GAT Asp	ACI Thi		T AC	G AA Ir As	C CC	G GA O Gl	u Gi	AA G(lu G)	GA G ly A	AC sp	TTG Leu	AAC Asr	ı Pr	G C	CG C	CG ro	GAA Glu		192
GCA Ala 65	. Ly a	G CA	G GT n Va	G CC 1 Pr	o va	T TC l Se 0	A TA	C TA	AC G	AT sp	TCA Ser 75	ACC	TA	T CT	TG A	GC er	ACA Thr 80		240
Asp	ASI		и шу	8 8	р ж s 5	C TAC	r Le	и гу	rs G.	90 17 /	Val	Thr	Ly	s Le	u P	ne 95	Glu		288
Arg	116	ту.	100	0	L AS	C CTO p Let	1 GI	y Ar 10	9 Me	et I	Seu	Leu	Thi	Se 11	r I]	le	Val		336
Arg	GIY	115	5 Pro	o Pne	e irī	G GGT p Gly	120	y Se O	r Tr	ır I	lle	Asp	Th:	Gl	u Le	u :	Lys		384
Val	130	ASL	Inz	ASI	ı cys	T ATT	ASI	ı va	1 11	e G	ln	Pro 140	Asp	Gl _j	y Se	r :	Tyr		432
145	Sei	GIU	GIU	Leu	150		vaı	LIL	≘ Il	e G 1	ly 5 5	Pro	Ser	Ala	a As	p]	[le [60	••	480
iie '	GIU	Pne	GIU	165	ьys	AGC Ser	Phe	: Gly	7 Hi 17	s G O	lu '	Val	Leu	Asr	1 Let	ս 1 5	Chr		528
Arg a	ASII	GIY	180	GIY	ser		Gin	185	Ile	e A:	rg 1	Phe	Ser	Pro 190	Ası	P	he		576 ·
ACG 1	rne	195	Pne	GIU	GIU	Ser	200	Glu	. Val	L·Aε	r q	Chr .	Asn 205	Pro	Leu	L	eu		624
	210	GIÀ	rys	Pne	Ala	215	Asp	Pro	Ala	\ Va	1 7	hr 1	Leu	Ala	His	G.	lu		672
CTG F Leu I 225	LIE :	nis	Ата	GIĀ	230	Arg	Leu	Tyr	Gly	23	e A 5	la 1	Ile	Asn	Pro	A: 24	sn 10		720
CGC G Arg V	aı i	rne	rys	245	Asn	Thr	Asn	Ala	Tyr 250	Ту	r G	lu M	1et	Ser	Gly 25 5	Le	eu.		768
GAA G Glu V	TA A	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GG Gl	T G y G	GC C ly H	lis .	GAT Asp 270	GCG Ala	A# Ly	AG 's		816
TTT A	Te t	SAC Asp 275	AGC Ser	TTG Leu	CAG Gln	Glu.	AAC Asn 280	GAG Glu	TTC Phe	CG	T C	eu T	AC ' Yr ' 85	TAC Tyr	TAC Tyr	AA As	C in		864
AAG T Lys P	TT A he I 90	AYA	GAT Asp	ATT Ile	Ala	AGT A Ser 1 295	ACA Thr	CTG Leu	AAC Asn	AAC Lys	G G(S A) 3 (la L	AG ? ys \$	rcc Ser	ATT Ile	GT Va	G 1		912

															-	
GG G1: 30:	7	C AC	r GCT r Ala	TCA Sea	Let 310	r GTL	TAT	T ATO	AA/ Lys	A AA: S Asi 319	n Va	T TT	T AA e Ly	A GA s Gl	G AAA u Lys 320	960
TA:	r CTC	CTA Lev	A TCT 1 Ser	GAA Glu 325	ASP	ACA Thr	TCT Ser	GG# Gly	AAA Lys	: Phe	TCC Se:	G GT	A GA	T AA. p Ly. 33:	A TTA s Leu 5	1008
AA) Lys	A TTI S Phe	GAT Asp	AAG Lys 340	TIEU	TAC	AAA Lys	ATG Met	TTA Leu 345	Thr	GAC Glu	AT:	TAC Tyl	C AC	r Gl	G GAT	1056
AA7 Asr	TTT Phe	GTI Val	. цуб	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	Leu	AAC Asn	AGA Arg	AAA Lys	A ACA 5 Thr 365	Ty	TTC Lev	AAT Asn	1104
TTT	GAT Asp 370	,	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro) Lys	GT#	AAT Asr	TAC Tyr	1152
ACA Thr 385	116	TAT	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AA1 Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415		1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
ATC Ile	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	GAT Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
TTT Phe	AAT Asn	TTT Phe 515	GAT Asp	AAT Asn	GAA Glu	Pro	GAA Glu 520	TAA neA	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	GIR	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632
CCT Pro 545	AAT Asn	GGA Gly	AAA Lys	AAG Lys	TAT Tyr 550	GAG Glu	TTA Leu	GAT Asp	AAA Lys	TAT Tyr 555	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 560	1680
CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 565	TTT Phe	GAA Glu	CAT His	Gly	AAA Lys 570	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr	1728

		'		5.80	GAA Glu	710	beu	neu	58	n Pi 5	ro s	er.	Arg	Val	Ту 59	T O	hr	Phe	1776
		5	95	υp	TAT Tyr	vai .	uys	600	va.	l As	n L	ys <i>l</i>	Ala	Thr 605	Gŀ	u A.	la	Ala	1824
	6	LO		· - ·	rgg ((515	GIII	net	ı va	1 1	yr F	15p 520	Phe	Th:	r As	sp	Glu	1872
62	5		•	u. .		30	. 1112	Asp	Lys	11.	е А. 6:	1a A	rsb	Ile	Thi	r Il	.e	Ile 640	1920
		0 .,	~ 1	6	GA C Sly F 45	10 4	ııa .	Leu	ASN	650	≥ G1	ГУ А	sn i	Met	Leu	1 Ty 65	r :	Lys	1968
		P - 1.	66	50	GT G	ıa D	eu .	rre	665	. Ser	r GI	у А	la (/al	Ile 670	Le	u I	Leu	2016
	· ·	67	5	.0 0	AG A lu I	TC Y	6	80	PIO	Val	. Le	u G	ly 1 6	hr 85	Phe	Ala	a I	eu	2064
	690	5 - 7		.c A.	CG A	69	95	al .	Leu	Thr	Va	1 G1	Ln T	hr :	Ile	Asp	A	sn	2112
705			. Ду	3 A.	GA AA cg As 71	.0	. u. j.	ys .	rrp	Asp	G1: 71!	ı Va 5	l T	yr I	Lys	Туг	7	le 20	2160
			• ••	72		а шу	'S V.	aı A	AST	730	Glr	ıIl	e As	sp L	eu	Ile 735	A:	rg	2208
-,-	272		740	5	A GC u Al	a ne	u G.	7	45	GIn	Ala	Gl	u Al	la T 7	hr 50	Lys	Α.	la	2256
	.10	755	. IYI	GI	G TA n Ty	r As	76	in T	yr '	Thr	Glu	Gl	u G1 76	lu L	λa	Asn	As	sn	2304
	770		ASI	•	T GA e As	77	5 16	u S	er:	ser	Lys	780	ı As	n G	lu .	Ser	Il	.e	2352
785	<i>D</i> , 5	714	nec		T AA: e As: 790)	: AS	on L	ys i	rne	Leu 795	Asr	ı Gl	n Cy	ys S	Ser	Va 80	0	2400
301	LYL	Deu	MEC.	805		Met	. 11	e Pi	ro 1	10 (10)	Gly	Val	. Ly	s Ar	rg I	Leu 315	Gl	u	2448
GAT Asp	rne	Asp	820	561	Let	. Lys	AS	9 AJ 82	1a 1. 25	eu 1	Leu	Lys	Ty:	r Il 83	e 1	yr	As	Þ	2496
AAT . Asn .	AGA Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT	GGT Gly	G1: 84:	n va	TA G	AT A	AGA Arg	TTA Leu	AAA Lys 849	s As	T A	AA 'ys	GT: Va:	r	2544

AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA ASN ASN Thr Leu Ser Thr ASP Ile Pro Phe Gln Leu Ser Lys Tyr Val

GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA ASP ASN Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys +

850

- (2) INFORMATION FOR SEQ ID NO: 12:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 879 amino acids

(B) TYPE: amino acid (D) TOPOLOGY: linear (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12: Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 235

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 475 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 490 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 520 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala

- Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 620
- Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 640
- Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 645 650 655
- Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670
- Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685
- Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700
- Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 705 710 715 720
- Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735
- Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750
- Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765
- Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 780
- Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800
- Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815
- Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830
- Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845
- Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
- Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 865 870 875
- (2) INFORMATION FOR SEQ ID NO: 13:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2862 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2862
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

AT Me	G CA t Gl	G T	TC G he V	TG A	AC A sn L 5	AG C ys G	AG T	rc A/ ne As	sn T	AT A yr I 10	AG (GAC Asp	CCT Pro	GT.	l As	AC sn	GGT Gly	48
GT Va	T GA l As	C A' p I		CC T la T 20	AC A' yr I	TC AM	AA AI /s Il	e Pr	CA A.	AC G	CC (GC Sly	CAG Gln	ATO Met	G1	AG .n	CCG Pro	96
GT(Val	G AA L Ly	3 .A.	TT T la Pi	rc A	AG A: ys I:	TT CA Le Hi	S AS	C AA n Ly 0	A AT	rc r Le r	GG C rp V	TT . 'al	ATT Ile 45	CCC	G GA	A u	CGC Arg	144
GAT Asp	Th:	L PI	T AG	CG AJ	AC CO	G GA O Gl	A GA u Gl 5	A GG u Gl	A GA y As	AC T	eu A	AC (sn)	CCG Pro	CCG	CC Pr	G (GAA Glu	192
GCA Ala 65	. Lys	G CA s Gl	G G1 n Va	G CC	.u va	T TC 1 Se	A TA	C TAC	C GA	p Se	CA A er T	CC 1	FAT Fyr	CTG Leu	AG Se:	C A	ACA Thr 80	240
GAC Asp	AA(Asr	GA 1 Gl	G AA u Ly	5 A5	T AA p As 5	C TA	C CTO	J Lys	G GG S G1 9	y Va	G A	CC A	ys Lys	TTA Leu	TTO Pho	2 (SAG Slu	288
CGT Arg	ATT	TA'	T TC r Se 10	Lin	T GA	C CTO	G GG(C CG1 / Arg 105	g Me	G CI t Le	G Ci u Le	rg A eu T	hr.	TCA Ser 110	ATC Ile	C G	TC al	336
CGC Arg	GGA Gly	ATO 11e	e PL	A TT o Ph	T TGG e Trj	G GG1 p Gly	GGC Gly	Ser	Th	C AT	T GA	p T	CG (hr (25	Glu	TTC	; A	AG ys	384
GTT Val	ATT Ile 130	vaf	C AC	r Ası	C TGC	C ATT	ASD	GTG Val	ATO	C CA Gl:	A CC n Pr 14	O A	AC (GGT Gly	AGC Ser	T	AC yr	432
AGA Arg 145	TCT Ser	GAA Glu	GAZ Glu	l Lei	AAC 1 Asr 150	CTC Leu	GTA Val	ATC Ile	ATO	GG(Gl)	y Pr	C To	CC C	GCG Ala	GAC Asp	I.	TT le 50	480
ATC Ile	CAG Gln	TTT Phe	GAC Glu	TGC Cys 165	. Lys	AGC Ser	TTT Phe	GGC Gly	CAC His	Gli	A GT 1 Va	G TI	IG A	sn	CTG Leu 175	A(Tì	CG nr	528
CGT Arg	AAC Asn	GIY	TAC Tyr 180	GIY	TCT Ser	ACT Thr	CAG Gln	Tyr	ATT Ile	CGT	TTO J Pho	C AC	r P	CA (ro .	GAC Asp	TT	C ne	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thi	AA As 20	n P	CG (CTG Leu	TI	G u	624
GGT (GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thi	: Le	G G	CA (CAC	GA Gl	.G u	672
CTG / Leu : 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCC Ala	AT Il	T A	AC C	Pro	AA As: 24	n	720
CGC (Arg (GTG /al	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	Ala	TAC Tyr 250	TAC Tyr	GAG Glu	AT(G AC	er G	GT 1y 55	TT: Le:	A. LL	768
GAA G Glu V	TA /	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CA:	T GA S As 27	A q	CG . la	AA(Lys	3	816

TT	T ATO	C GAG Asp 279	, ,,,	C TTO	G CAC	GAG Glu	AAC Asn 280	GIL	TTO Phe	C CG	r cro	TA:	r Ty	C TA	C AAC	864
AA Ly	G TTT s Phe 290	- 27 C	GA1 Asp	ATI Ile	GCA Ala	AGT Ser 295	Int	Leu	AAC Asn	Lys	GCT Ala 300	a Lys	TC(C AT	T GTG ≥ Val	912
30	5		Ala	ser	310	Gin	Tyr	Met	Lys	315	ı Val	. Phe	Lys	5 Glu	AAA Lys 320	960
Ly.	L Dec	. neu	Je!	325	Asp	inr	ser	GIY	330	Phe	Ser	· Val	Asp	335		1008
		vab	340	Deu	TYL	Lys	мес	345	Thr	Glu	Ile	Tyr	350	Glu	GAT Asp	1056
ASI	. FIIC	355	пуз	· Pne	Pne	Lys	360	Leu	Asn	Arg	Lys	Thr 365	Tyr	Leu	AAT Asn	1104
7110	GAT Asp 370	Буб	Ala	vai	Pne	375	IIe	Asn	Ile	Val	9ro 380	Lys	Val	Asn	Tyr	1152
385		TYL	Asp	GIÀ	390	ASN	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	Asn 400	1200
	AAT Asn	GTÅ	GIII	405	Int	GIU	He	Asn	410	Met	Asn	Phe	Thr	Lys 415	Leu	1248
Lys	AAT Asn	FIIE	420	GIY	Leu	Pne	GIu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	Arg	1296
GCG	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
ATC	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT . Ile '	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA (Leu .	Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
TTT Phe	AAT Asn	TTT Phe 515	GAT Asp	AAT Asn	GAA Glu	Pro (GAA A Glu A 520	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC (Gin	TTA (Leu (535	GAA (Glu)	CTT Leu	ATG Met	Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632

545	., bys bys	550	wab riva	TAT ACT ATG	Phe His Tyr 560	1680
	565	me did nis	570	TCT AGG ATT (Ser Arg Ile A	la Leu Thr 575	1728
AAT TCT GT Asn Ser Va	T AAC GAA (l Asn Glu / 580	GCA TTA TTA La Leu Leu	AAT CCT I Asn Pro S	AGT CGT GTT T Ser Arg Val T	AT ACA TTT Yr Thr Phe 90	1776
599	5	600	vai Asn I	AAA GCT ACG G Lys Ala Thr G 605	lu Ala Ala	1824
610	- G1, 11p (615	Leu vai 1	TAT GAT TTT A Tyr Asp Phe T 620	hr Asp Glu	1872
625	6	30.	rys ite A	CCG GAT ATA AG la Asp Ile TI 35	ar Ile Ile 640	1920
110 110 191	645	to Ala Leu	Asn Ile G 650	GT AAT ATG TT ly Asn Met Le	eu Tyr Lys 655	1968
	660 A.	a Leu Ile	ene ser G. 665	GA GCT GTT AT ly Ala Val II 67	e Leu Leu 0	2016
675	110 G14 11	680	Pro Val Le	TA GGT ACT TT eu Gly Thr Ph 685	e Ala Leu	2064
GTA TCA TAT Val Ser Tyr 690	ATT GCG AA Ile Ala As	T AAG GTT (n Lys Val I 695	TA ACC GT Leu Thr Va	TT CAA ACA AT il Gln Thr Il 700	A GAT AAT e Asp Asn	2112
GCT TTA AGT Ala Leu Ser 705	AAA AGA AA Lys Arg As 71	n Gra rae t	GG GAT GA TP Asp Gl 71	G GTC TAT AA. u Val Tyr Ly: 5	A TAT ATA 5 Tyr Ile 720	2160
GTA ACA AAT Val Thr Asn	TIP DEG AL	A AAG GTT A a Lys Val A	sn Inr GI	G ATT GAT CTA n Ile Asp Lev	A ATA AGA 1 Ile Arg 735	2208
בין טיין דור בין בי	AAA GAA GC Lys Glu Ala 740	Ped Gld W	AT CAA GC sn Gln Al 45	A GAA GCA ACA a Glu Ala Thr 750	Lys Ala	2256
ATA ATA AAC Ile Ile Asn 755	TAT CAG TAT Tyr Gln Tyr	AAT CAA T Asn Gln T 760	AT ACT GAO	G GAA GAG AAA 1 Glu Glu Lys 765	AAT AAT Asn Asn	2304
ATT AAT TTT I Ile Asn Phe I 770	AAT ATT GAT Asn Ile Asp	GAT TTA AG Asp Leu Se 775	GT TCG AAA er Ser Lys	A CTT AAT GAG Leu Asn Glu 780	TCT ATA Ser Ile	2352
AAT AAA GCT A Asn Lys Ala N 785	ATG ATT AAT Met Ile Asn 790	ATA AAT AA Ile Asn Ly	AA TTT TTO s Phe Leu 795	AAT CAA TGC Asn Gln Cys	TCT GTT Ser Val 800	2400
TCA TAT TTA A Ser Tyr Leu M	ATG AAT TCT Met Asn Ser 805	ATG ATC CO	T TAT GGT TO Tyr Gly 810	GTT AAA CGG Val Lys Arg	TTA GAA Leu Glu 815	2448

GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT Tyr	GAT Asp		2496
AAT Asn	AGA Arg	GGA Gly 835	TIIT	TTA Leu	ATT	GGT Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val		2544
AAT Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val		2592
GAT Asp 865	AAT Asn	CAÁ Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880		2640
CCT Pro	GGA Gly	CCG Pro	GAG Glu	ACG Thr 885	CTC Leu	TGC Cys	GGG Gly	GCT Ala	GAG Glu 890	CTG Leu	GTG Val	GAT Asp	GCT Ala	CTT Leu 895	CAG Gln		2688
TTC Phe	GTG Val	TGT Cys	GGA Gly 900	GAC Asp	AGG Arg	GGC Gly	TTT Phe	TAT Tyr 905	TTC Phe	AAC Asn	AAG Lys	CCC Pro	ACA Thr 910	GGG Gly	TAT Tyr		2736
GGC Gly	TCC Ser	AGC Ser 915	AGT Ser	CGG Arg	AGG Arg	GCG Ala	CCT Pro 920	CAG Gln	ACA Thr	GGT Gly	ATC Ile	GTG Val 925	GAT Asp	GAG Glu	TGC Cys		2784
TGC Cys	TTC Phe 930	CGG Arg	AGC Ser	TGT Cys	GAT Asp	CTA Leu 935	AGG Arg	AGG Arg	CTG Leu	GAG Glu	ATG Met 940	TAT Tyr	TGC Cys	GCA Ala	CCC Pro		2832
CTC Leu 945	AAG Lys	CCT Pro	GCC Ala	AAG Lys	TCA Ser 950	GCT Ala	GAA Glu	GCT Ala	TAG	•	Í					:	2862

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 954 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110

- Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
- Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140
- Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160
- Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175
- Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190
- Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205
- Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220
- Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 230 235
- Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255
- Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270
- Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285
- Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300
- Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320
- Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335
- Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350
- Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365
- Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 380
- Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400
- Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415
- Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
- Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435
 440
 445
- Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460

Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480

Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495

Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510

Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525

Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 540

Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560

Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575

Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590

Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 620

Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640

Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650 655

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
725 730 735

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 780

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815

288

As	p Ph	e As	p Al 82	a Se O	r Le	u Lys	s Asp	Ala 825	Lev	ı Lev	Lys	Ty	r Il 83		yr	Asp		
As	n Ar	g Gl 83	y Th	r Le	u Il	e Gly	/ Glr 840	Val	Asp	Arg	Leu	Ly:		p Ly	/s	Val		
As	n Asi	n Th:	r Le	u Se	r Th	r Asp 855	Ile	Pro	Phe	Gln	Leu 860	Sei	r Ly	s Ty	/r	Val		
As 86	p Ası 5	n Gli	n Arg	g Lei	1 Let 870	ı Ser	Thr	Phe	Thr	Glu 875	Tyr	Ile	≥ Ly	s Se		Arg 880		
Pr	o Gly	y Pro	Glu	1 Thi 885	r Leu	ı Cys	Gly	Ala	Glu 890	Leu	Val	Asp	Al.	a Le 89		Gln		
Pho	e Val	l Cys	900	/ Asp	Arg	g Gly	Phe	Tyr 905	Phe	Asn	Lys	Pro	910		у '	Tyr		
Gly	/ Ser	Ser 915	Ser	Arg	, Arg	, Ala	Pro 920	Gln	Thr	Gly	Ile	Val 925	Ası	Gl	u (Cys		
Cys	930	Arg	Ser	. Cys	Asp	Leu 935	Arg	Arg	Leu	Glu	Met 940	Tyr	Cys	Al.	a I	Pro		
Leu 945		Pro	Ala	Lys	Ser 950	Ala	Glu	Ala	*				•					
(2)	INF	ORMA	TION	FOR	SEQ	ID N	10: 1	.5 :										
	(ii)	() () () () MOI () FEX	A) Li B) T' C) S' C) TO LECUI ATURE A) NA	ENGTI YPE: TRAMI OPOLO LE TI E: AME/I	H: 2 nuc DEDNI OGY: YPE:	CTERITION OF THE PROPERTY OF T	ease acid doub ar (gen	pair l										
						PTIO												
Met 1	Gln	Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	rne I	Asn 3	yr I	AAG C Lys A	ا sp	CCT Pro	Val	AAC Asn 15	G.	GT Ly		48
GTT Val	GAC Asp	ATT Ile	GCC Ala 20	TAC Tyr	ATC Ile	AAA . Lys	ATT (CCA A Pro A 25	AC C	GCC G Ala G	GC (CAG . Gln !	ATG Met 30	CAG Gln	CC	CG CO		96
GTG Val	AAG Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT Ile	CAT /	AAC A Asn I 40	AAA A Lys I	TC T	GG G	TT A	le i	CCG Pro	GAA Glu	CG	G G	1	44
GAT Asp	ACA Thr 50	TTT . Phe	ACG . Thr .	AAC Asn	CCG (Pro	GAA (Glu (55	GAA G	GA G	AC T	eu A	AC C sn P 60	CG (CCG Pro	CCG Pro	GA Gl	A u	1	92
GCA Ala 65	AAG (CAG (Gln '	GTG (CCA (GTT 'Val :	TCA 1 Ser 1	CAC I	AC G	sp S	CA A er T	CC T hr T	'AT (CTG Leu	AGC Ser	Th	A	24	40

GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85

		•	10	0		, 100	. 01	10	9 Me 5	c re	u Le	u Th	r Se 11	r Il O	C GTC e Val	336
•		11:	5			July	120)	r rin	r 116	e As	P Th:	r Gl	u Le	G AAG u Lys	384
	130)			. cys	135	ASI	ı val	116	GIr	1 Pro	o Asp	Gl;	y Se	C TAC r Tyr	432
145					150	Deu	Val	ıre	: 116	155	Pro	Ser	Ala	a Ası	C ATT P Ile 160	480
				165	Lys	Jer	PHE	GIŸ	170	Glu	Val	. Leu	Asr	1 Let 175	-	528
J		1	180	O1,	Der	1111	GIN	185	ııe	Arg	Phe	Ser	Pro 190	Ast	TTC Phe	576
	-,	195		Jiu	Jiu	361	200	GIU	vai	Asp	Thr	Asn 205	Pro	Leu	TTG Leu	624
2	210	J=7	5 75	1110	ALG	215	Asp	PIO	ATA	Val	220	Leu	Ala	His	GAG Glu	672
225			7124	Gly	230	Arg	Leu	lyr	GIY	235	Ala	Ile	Asn	Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG . Glu .	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	WIG	AGT A Ser ' 295	ACA Thr	CTG Leu	AAC Asn	Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	ser	TTA Leu 310	CAG : Gln :	rat . Fyr	ATG Met	Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	Ser	GAA (Glu / 325	GAT / Asp '	ACA :	CT :	GIY	AAA Lys 330	TTT 'Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA 1 Leu 1	TAC A	AAA A Lys N	tet :	TTA . Leu '	ACA (GAG A	ATT Ile	Tyr '	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT ' Asn	FIIE	GTT Val 355	AAG ' Lys	TTT :	TTT / Phe 1	ras /	GTA (Val 1 860	CTT . Leu .	AAC A	AGA A	Lys '	ACA ' Thr ' 365	TAT Tyr	TTG Leu	AAT Asn	1104

	370	273	7.2	· vai	. File	375	; ;	e As	n 11	le Va	11 P:	ro Ly BO	ys V	al A	sn	TAC Tyr		1152
385	110	.,.	лэр	Gly	TTT Phe 390	ASII	Let	ı Ar	g As	n Th	r As	sn Le	eu A	la A	la	Asn 400		1200
FIIC	ASII	Gly	GIII	405	ACA Thr	GIU	116	: Ası	n As 41	n Me O	t As	n Ph	e Ti	1r L	ys 15	Leu		1248
273	*15.:	1110	420	GIY	TTG Leu	Pne	GIU	429	Y TY	r Ly	s Le	u Le	u Cy 43	s Va 0	al	Arg		1296
GGG Gly		435	****	Jer	Lys	1111	440	ser	Let	ı Ası	p Ly	s G1 44	у Ту 5	r As	in	Lys		1344
	450	Gly	Arg	Cys	Asp	455	Ата	Leu	. Asr	ı Asp	46	ц Су: 0	s Il	e Ly	's	Val		1392
AAT A Asn A 465		· q··	rap	neu	470	Pne	ser	Pro	Ser	475	Ası) Ası	ı Ph	e Th	r 1	Asn 180		1440
GAT (Asp I	Jeu ,	1511 1	Lys '	485	GIU	GIU	11e	Thr	Ser 490	Asp	Thi	: Asr	ı Ile	e G1:	u 7 5	Ala		1488
GCA G Ala G	, Lu C	5	500	116 .	ser .	Leu .	Asp	505	11e	Gln	Gln	Tyr	Ty:	Lei	ı T	'hr		1536
TTT A Phe A	5	15	sp.	ASII (JIU 1	Pro (520	Asn	He	Ser	Ile	Glu 525	Asn	Le.	ıS	er		1584
	30	16 1	16 0	sry C	5111	35 35	31U	Leu	Met	Pro	Asn 540	Ile	Glu	Arg	P	he		1632
CCT A Pro A 545	J., G.	-y L	ys L	,ys 1 5	50	ilu L	eu .	Asp	Lys	Tyr 555	Thr	Met	Phe	His	T:	yr 50		1680
CTT CC Leu Ai	.g A.	La G.	5	65	ne G	iu n	115 (ìΙλ	Lys 570	Ser	Arg	Ile	Ala	Leu 575	Tì	ır		1728
AAT TO Asn Se	IL V	5 E	90	IU A	Ta P	eu L	eu A	185 185	Pro	Ser	Arg	Val	Tyr 590	Thr	Pł	ıe	:	1776
TTT TO Phe Se	T TO F Se 59	T W	AC T	AT G yr V	TA A al L	ys Ի	AA C ys V 00	TT I	AAT Asn	AAA (Lys .	GCT Ala	ACG Thr 605	GAG Glu	GCA Ala	GC Al	T .a	1	L824
ATG TT Met Ph 61	ie ne	CA GC	SC TO Ly Ti	GG G: pp Va	al G	AA C lu G 15	AA T ln L	TA (GTA ' Val '	Tyr I	GAT Asp 620	TTT Phe	ACC Thr	GAT Asp	GA Gl	A u	1	.872
ACT AG Thr Se 625	C GA r Gl	A GT u Va	TA AC	er 11	OT AC	CG G	AT A sp L	AA A ys 1	le A	GCG (Ala # 635	GAT Asp	ATA Ile	ACT Thr	ATA Ile	AT Il 64	e	. 1	920

AT	T CCI e Pro	A TAT	T ATA	GG# Gly 645	PEC	GCT Ala	TTA Leu	AA1 Asr	ATA 11e 650	: Gly	AAT Asn	C ATO	TTI Let	A TA' 1 Ty: 65:	T AAA r Lys		1968
GA' Asj	T GAT	TTT Phe	GTA Val 660	GIA	GCT Ala	TTA Leu	ATA Ile	Phe	Ser	GGA Gly	GCT Ala	GTI Val	ATT Ile	Let	TTA Leu		2016
GAI Glu	A TTI 1 Phe	11e 675	Pro	GAG Glu	ATT	GCA Ala	ATA Ile 680	Pro	GTA Val	TTA Leu	GGT	ACT Thr 685	Phe	GCA Ala	CTT Leu		2064
GT/ Val	TCA Ser 690	TYT	ATT Ile	GCG Ala	AAT A sn	AAG Lys 695	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	Thr	ATA	GAT Asp	AAT Asn		2112
GCT Ala 705	Leu	AGT	AAA Lys	AGA Arg	AAT Asn 710	GAA Glu	AAA Lys	TGG	GAT Asp	GAG Glu 715	GTC Val	TAT Tyr	AAA Lys	TAT	ATA Ile 720	. •	2160
GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg		2208
AAA Lys	. AAA Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala		2256
ATA Ile	ATA Ile	AAC Asn 755	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AAT Asn		2304
ATT Ile	AAT Asn 770	TTT Phe	AAT Asn	ATT	GAT Asp	GAT Asp 775	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	AAT Asn	GAG Glu	TCT Ser	ATA Ile		2352
AAT Asn 785	AAA Lys	GCT Ala	ATG Met	ATT Ile	AAT Asn 790	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 795	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 800		2400
TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GAA Glu		2448
GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 630	TAT Tyr	GAT Asp		2496
AAT Asn	AGA Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT Ile	Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val		2544
AAT Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val		2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880		2640
CCT Pro	CAA Gln	TCT Ser	AAA Lys	GTT Val 885	AAA Lys	AGA Arg	CAA . Gln	Ile	TTT Phe 890	TCA Ser	GGC Gly	TAT Tyr	Gln	TCT Ser 895	GAT Asp		2688
ATT Ile	GAT Asp	ACA Thr	CAT . His . 900	AAT Asn	AGA Arg	ATT . Ile	Lys .	GAT Asp 905	GAA Glu	TTA Leu	TGA *						2724

- (2) INFORMATION FOR SEQ ID NO: 16:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 908 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 470 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 520 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 635 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys

- Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670
- Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685
- Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700
- Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720
- Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735
- Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750
- Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765
- Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 780
- Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800
- Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815
- Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830
- Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845
- Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860
- Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg 865 870 875 880
- Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp 885 890 895
- Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu
 900 905
- (2) INFORMATION FOR SEQ ID NO: 17:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3042 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..3042
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1

GT1 Val	GAC L Asp	ATT	GCC Ala 20	r ry	T ATO	AAA Lys	ATI	CCF Pro) Asr	GCC Ala	GGG Gly	C CAC	ATO Met	Glr	G CCG 1 Pro	96
GTC Val	AAC Lys	GCT Ala 35	Pne	Lys	ATT	CAT His	AAC Asn 40	Lys	ATC	TGG	GTT Val	ATI Ile 45	Pro	GAZ Glu	CGC Arg	144
GAT Asp	ACA Thr	FILE	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	Pro	CCG	CCG	GAA Glu	192
GCA Ala 65	Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	ser	TAC Tyr	TAC	GAT Asp	TCA Ser 75	ACC Thr	TAT	CTG Leu	AGC Ser	ACA Thr 80	240
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC Tyr	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC Thr	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu	288
CGT Arg	ATT	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	· 672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	Glu	AAC Asn 280	Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864

A)	AG 1 Ys P	TT I	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTC	AA As	C Al	s A	CT A	AAG :	rcc . Ser :	ATT Ile	GTG Val	912
G(G1 30	GT A ly T 05	CC A	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AA Ly	A AA S As	n V	TT 1 al P	TT #	AAA (GAG Glu	AAA Lys 320	960
,					GAA (Glu / 325	rsp .		ser	GIY	330	s Ph O	e Se	er V	al A	sp L	ys 35	Leu	1008
-2				340	TTA 1 Leu 1	. 7 - 1	Jys I	MEC	345	ini	r Gl	u II	le T	yr T 3	hr G 50	lu	Asp	1056
		3 :	55	., -	rrr 1 Phe E		:	360	reu	Asn	1 Arg	g Ly	rs Th	nr Ty 55	yr L	eu	Asn	1104
•	37	o -		'	TA T	3	75	.IE	ASII.	TIE	· val	38	O LY	/S Vā	al A:	sn	Tyr	1152
385	5	1	- ••	JP C		90	311 L	eu /	arg	Asn	395	As	n Le	u Al	a Al	la .	Asn 400	1200
			, C.	4	AT A sn T 05	G.	Lu I	Te Y	ASN .	410	Met	Ası	n Ph	e Th	r Ly 41	rs 1 .5	Leu	1248
- 2 -			4.2	20	GA T' ly Le	iu Pi	ie G	1 U F	25	ryr	Lys	Leu	ı Le	u Cy 43	s Va O	1 4	lrg	1296
		43	5		CT AA ≥r Ly		4.4	10	erı	ren	Asp	Lys	Gly	y Ty:	r As:	n L	ys	1344
	450			9 0	GC GA 's As	45	5 5	ابل ده.	eu A	isn	Asp	Leu 460	Cys	: Ile	Ly:	s V	al	1392
465				ם בכ	G TT u Ph 47	0	- 56	I P	ro s	er	G1u 475	Asp	Asn	Phe	Thi	4 (sn 80	1440
•			,	48		u GI	4 11	e 11	4	er / 90	Asp	Thr	Asn	Ile	Glu 495	A.	la	1488
			500)	T AG' e Sei	. Det	ı AS	50 50	u 1.	re (GIn (Gln	Tyr	Ту <u>г</u> 510	Leu	Ţŀ	ır	1536
		515		, AS	r GA/ n Glu	PIC	520)	n I.	Le S	ser :	Ile	Glu 525	Asn	Leu	Se	r	1584
	530		110	. G.L.;	, GII.	535	GI	те	u Me	et P	ro A	4sn 540	Ile	Glu	Arg	Ph	е	1632
CCT / Pro / 545	AAT Asn	GGA Gly	AAA Lys	Lys	TAT Tyr 550	GIU	TTA Lev	A GA' 1 As _l	T AA p Ly	'S T	AT A Yr 1 55	CT	ATG Met	TTC Phe	CAT His	TA Ty 56	r	1680

				56	5		- 111.	3 G1	57	s se O	r Ar	g 11	e Al	a Le 57		r
			58	0			ı net	589	5	5 Se	r Ar	g Va	1 Ty 59	r Th	A TT	е
		59	5	F - 1	_ , ,	ı Dya	600	va.	L ASI	л гуз	S Al	a Th 60	r Gl	u Al	A GC a Ala	a
	61	ם				615	GII	L Let	ı vaı	Туг	620	Phe	≘ Th:	r As	T GAZ P Glu	1
625	5			- 00.	630)	Asp	rys	ille	635	Asp) Ile	≥ Thi	: Il	A ATT e Ile 640	
		-,		645	,	, ,,	neu	ASII	650	GIY	Asn	Met	Leu	Ty:	_	•
•			660)	7124	Dea	116	665	ser	GIÀ	Ala	Val	11e	Let	G TTA Leu	
		675	5	014		ALG	680	PIO	vai	Leu	Gly	Thr 685	Phe	Ala	CTT Leu	2064
	690	-1-			AJII	695	val	ren	inr	Vai	G1n 700	Thr	Ile	Asp	AAT Asn	2112
705			a, s	Arg	710	Giu	пуѕ	irp	Asp	715	Val	Tyr	Lys	Tyr	720	2160
			TGG Trp	725	AIG	. uys	vai	ASN	730	Gin	Ile	Asp	Leu	Ile 735	Arg	2208
-1-	-75		AAA Lys 740	, JII	AIA	Deu	GIU	745	Gin	Ala	Glu	Ala	Thr 750	Lys	Ala	2256
	÷c	755	TAT Tyr	GIII	ıyı	ASN	760	Tyr	Thr	Glu	Glu	Glu 765	Lys	Asn	Asn	2304
***	770	rne	AAT Asn	116	Азр	775	Leu .	ser	Ser	Lys	Leu 780	Asn	Glu	Ser	Ile	2352
785	Буз	AIG	ATG Met	116	790	ile .	Asn)	Lys	Phe	Leu . 795	Asn	Gln	Cys	Ser	Val 800	2400
361	IYL	neu	ATG Met	805	ser	мес	ile i	Pro '	Tyr (810	Gly '	Val	Lys	Arg	Leu 815	Glu	2448
GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT . Leu	AAA (Lys /	Asp A	GCA 1 Ala 1 B25	TTA '	TTA /	AAG Lys	Tyr	ATA Ile 830	TAT Tyr	GAT Asp	2496

AA:	r AG	A GG. g G1; 83:	y + 1.1	T TT. r Le	A AT	r GG7 e Gly	CAA Glr 840	ı va.	A GA1 l Asp	AGA Arg	TTA Lev	A AAA 1 Lys 845	: Ası	r aa o Ly:	A GTT s Val	2544
AAT Asr	AA 1 ASI 028		A CT	T AGʻ u Sei	r ACA	A GAT Asp 855) TIE	CCT Pro	TTT Phe	CAG Gln	CTI Leu 860	ı Ser	AAA Lys	А ТАС 5 Туз	C GTA C Val	2592
865	, Asi	GII	ı Arç	a rec	870	ser	Thr	Phe	Thr	Glu 875	Tyr	Ile	Lys	Ser	GGC Gly 880	2640
Deu	. vai	. 3e1	PIC	885	MIA	. Ата	HIS	Tyr	890	Gln	His	Asp	Glu	Ala 895		2688
GAC Asp	AAC Asn	AAA Lys	Phe 900	. Wan	AAA Lys	GAA Glu	CAA Gln	CAA Gln 905	AAC Asn	GCG Ala	TTC Phe	TAT	GAG Glu 910	ATC Ile	TTA Leu	2736
CAT His	TTA Leu	CCT Pro 915	ASII	TTA Leu	AAC Asn	GAA Glu	GAA Glu 920	CAA Gln	CGA Arg	AAC Asn	GCC Ala	TTC Phe 925	ATC Ile	CAA Gln	AGT Ser	2784
TTA Leu	AAA Lys 930	GAT Asp	GAC Asp	CCA Pro	AGC Ser	CAA Gln 935	AGC Ser	GCT. Ala	AAC Asn	CTT Leu	TTA Leu 940	GCA Ala	GAA Glu	GCT Ala	AAA Lys	2832
AAG Lys 945	CTA Leu	AAT Asn	GAT Asp	GCT Ala	CAG Gln 950	GCG Ala	CCG Pro	AAA Lys	GTA Val	GAC Asp 955	AAC Asn	AAA Lys	TTC Phe	AAC Asn	AAA Lys 960	2880
GAA Glu	CAA Gln	CAA Gln	AAC Asn	GCG Ala 965	TTC Phe	TAT Tyr	GAG Glu	ATC Ile	TTA Leu 970	CAT His	TTA Leu	CCT Pro	AAC Asn	TTA Leu 975	AAC Asn	2928
GAA Glu	GAA Glu	CAA Gln	CGA Arg 980	AAC Asn	GCC Ala	TTC Phe	Ile	CAA Gln 985	AGT Ser	TTA . Leu :	AAA Lys	Asp .	GAC Asp 990	CCA Pro	AGC Ser	2976
CAA Gln	AGC Ser	GCT Ala 995	AAC Asn	CTT Leu	TTA Leu	Ala	GAA (Glu / 1000	GCT Ala	AAA . Lys !	AAG (Lys 1	Leu ,	AAT (Asn)	GAT Asp	GCT Ala	CAG Gln	3024
lla	CCG Pro	Lys	GTA Val	GAC Asp	TAG *											3042

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1014 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 135 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 280 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 470 475 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 520 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 570 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 615 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 665 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 745

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
770 780

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly 865 870 875

Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val 885 890 895

Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu 900 905 910

His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 915 920 925

Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys 930 935 940

Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 945 950 955 960

Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Prò Asn Leu Asn 965 970 975

Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 980 985 990

Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000 1005

Ala Pro Lys Val Asp 1010

(2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3509 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..3509
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Me	1	. O V.	4.1. 11		5 5	AT AA sn As	n PN	e As	n Ty	r As 0	n As	p Pr	:0 I]	le A	sp A	sn	48
AA As	T AA n As	T AT	Le I	TT AT Le Me 20	TG AT	G GA	G CC u Pro	T CC o Pr 2	o Ph	T GC e Al	G AG a Ar	A GG g Gl	y Th	G G(F G]	GG AG	GA rg	96
TA'	r TA r Ty	r ry	A GO 's Al	T TI a Ph	T AA ie Ly	A ATO	C AC	r As	r CG	T AT	T TG	G AT. p Il. 4	e Il	A CC e Pr	G G/ O G1	lA .u	144
AG/ Arg	A TA 3 Ty 5	r in	T TI r Ph	T GG e Gl	A TA	T AAA r Lys 55	Pro	GA(G GAT	TTT Phe	T AA? Asi 60	1 Ly	A AG s Se	T TC r Se	C GC r Gl	T Y	192
ATT Ile 69	PIL	T AA e As	T AG n Ar	A GA g As	T GT p Val 70	r rgi l Cys)	GAA Glu	TA1	TAT	GAT Asp 75	Pro	A GAT	TAC Ty:	C TT r Le	u As	T n 0	240
ACI Thr	`AA: Asi	r GA	T AA p Ly	A AAG S Ly:	s Asr	T ATA	TTT Phe	TTA Leu	CAA Gln 90	Thr	ATC Met	ATC	AAC Lys	G TT.	u Ph	T e	288
AAT Asn	AG/ Arg	A ATO	C AA E Ly: 10	s Sei	A AAA r Lys	CCA Pro	TTG Leu	GGT Gly 105	Glu	AAG Lys	TTA Leu	TTA Leu	GAC Glu	Me	G AT	r e	336
ATA Ile	AAT Asr	GGT Gly 115	. 116	A CCT	TAT Tyr	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	Leu	GAZ Gli	A GAG	3	384
TTT Phe	AAC Asn 130	Thi	AA(Asr	ATI Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AG1 Ser	`AAT		432
CCA Pro 145	GGA Gly	GAA Glu	GTC Val	Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile	:	480
TTT Phe	GGA Gly	CC T Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly		528
ATA Ile	CAA Gln	AAT Asn	CAT His 180	TTT Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	Gly	TTC Phe	Gly	Gly	Ile	ATG Met	CAA Gln		576
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu		624
Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	ATA Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT Arg	Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro		672
GCC Ala 225	TTG Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAA (Glu)	CTT . Leu	ATA Ile	His	GTT (Val : 235	TTA Leu	CAT His	GGA Gly	TTA Leu	TAT Tyr 240		720
GGC Gly	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA (CCA A	Ile	GTA (Val 250	CCA /	AAT (Asn (GAA . Glu	Lys	AAA Lys 255	TTT Phe		768
TTT A	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT A	[le (CAG (Gln / 265	GCA (Ala (GAA (Glu (GAA (Glu)	Leu '	TAT Tyr 270	ACA Thr	TTT Phe		816

	•															
GG/ Gly	A GGA	Gln 275	Asp	Pro	AGC Ser	ATC	ATA Ile 280	Thr	Pro	TCT Ser	ACG Thr	GAT Asp 285	Lys	AGT Ser	ATC	864
TAT	GAT Asp 290	rys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	Asp	AGA Arg	CTT Leu	AAC Asn	912
AAC Lys 305	vai	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT	TTT	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	Glu	Asp	TTG Leu	Phe	TTT Phe	Ile	GCT Ala	Asp	Lys	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1632

AC Th 54	ır Ph	T CC	T CT o Le	A GA u As	T AT p Il 55	e Arg	A GAT g Asi	T AT	A AG e Se	T TI r Le 55	u Th	CA TO	T TO	CA TO	TT GAT ne Asp 560	> .	
GA As	T GC p Al	A TT a Le	A TT u Le	A TT u Pho 56	e Se	T AAC r Ası	C AAA n Lys	Val	T TA' 1 Ty: 57	r Se	A TT r Ph	T TT e Ph	T TO	T Aler Me	G GAI	1728	
Ту	r II	e Ly	5 Th	r Ala O	a Asr	ı Lys	. Val	Val 589	l Gli	ı Al	a Gl	y Le	ս Ph 59	e Al O	A GGT a Gly	,	
Tr	p Va	1 Ly 59	s Gli 5	n Ile	≥ Val	l Asn	Asp 600	Phe	e Val	l Ile	e Gl	u Al.	a As 5	n Ly	A AGC s Ser	;	
AA As:	T AC	r Mei	G GA:	r AAA	ATT	GCA Ala 615	Asp	ATA	TCT Ser	CT/	A AT 1 Ile 620	e Vai	r cc l Pr	T TA	T ATA	1872	
625	/ Leu	ı Ala	a Leu	ı Asn	630	Gly	Asn	Glu	Thr	635	Lys	s Gly	/ Ası	n Phe	GAA Glu 640	1920	
AA:	r GCT n Ala	TTT Phe	GAG Glu	Ile 645	Ala	GGA Gly	GCC Ala	AGT Ser	ATT Ile 650	Leu	CT/	A ĞAZ ı Glu	A TT	T ATA 116 659	A CCA Pro	1968	
GAA Glu	CTI Leu	TTA Leu	ATA Ile 660	Pro	GTA Val	GTT Val	GGA Gly	GCC Ala 665	TTT Phe	TTA Leu	TTA Leu	GAA Glu	TCA Sex 670	Ty	ATT Ile	2016	
Asp	Asn	Lys 675	Asn	Lys	Ile	Ile	Lys 680	Thr	Ile	Asp	Asn	Ala 685	Leu	Thr	Lys	2064	
AGA Arg	AAT Asn 690	GAA Glu	AAA Lys	TGG Trp	AGT Ser	GAT Asp 695	ATG Met	TAC Tyr	GGA Gly	TTA Leu	ATA Ile 700	GTA Val	GCG	CAA Gln	TGG Trp	2112	
CTC Leu 705	TCA Ser	ACA Thr	GTT Val	AAT Asn	ACT Thr 710	CAA Gln	TTT Phe	TAT Tyr	ACA Thr	ATA Ile 715	AAA Lys	GAG Glu	GGA Gly	ATĠ Met	TAT Tyr 720	2160	
AAG Lys	GCT Ala	TTA Leu	AAT Asn	TAT Tyr 725	CAA Gln	GCA Ala	CAA Gln	GCA Ala	TTG Leu 730	GAA Glu	GAA Glu	ATA Ile	ATA Ile	AAA Lys 735	TAC Tyr	2208	
AGA Arg	TAT Tyr	AAT Asn	ATA Ile 740	TAT Tyr	TCT Ser	GAA Glu	Lys	GAA Glu 745	AAG Lys	TCA Ser	AAT Asn	ATT Ile	AAC Asn 750	ATC Ile	GAT Asp	2256	
TTT Phe	AAT Asn	GAT Asp 755	ATA Ile	AAT Asn	TCT Ser	Lys	CTT . Leu . 760	AAT Asn	GAG Glu	GGT Gly	ATT	AAC Asn 765	CAA Gln	GCT Ala	ATA Ile	2304	
GAT Asp	AAT Asn 770	ATA Ile	AAT Asn	AAT Asn	TTT Phe	ATA Ile 775	AAT (Asn (GGA Gly	TGT Cys	TCT Ser	GTA Val 780	TCA Ser	TAT Tyr	TTA Leu	ATG Met	2352	
AAA Lys 785	AAA Lys	ATG Met	ATT Ile	CCA Pro	TTA Leu 790	GCT (Ala	GTA (Val (GAA . Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400	
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA I	AAT 1 Asn 1	Гуr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448	

Let ITE GIY	AGT GCA GAA TA Ser Ala Glu Ty 820	825	Ser Lys Val	Asn Lys Tyr 830	Leu
835	ATG CCG TTT GA Met Pro Phe As	840	lle Tyr Thr	Asn Asp Thr 845	Ile
CTA ATA GAA Leu Ile Glu 850	ATG TTT AAT AF Met Phe Asn Ly 85	s Tyr Asn	AGC GAA ATT Ser Glu Ile 860	TTA AAT AAT Leu Asn Asn	ATT 2592 Ile
865	TTA AGA TAT AA Leu Arg Tyr Ly 870	s Asp Asn A	Asn Leu Ile 875	Asp Leu Ser	Gly 880
TYP GIY ATA	AAG GTA GAG GT Lys Val Glu Va 885	I Tyr Asp C	Gly Val Glu 890	Leu Asn Asp 895	Lys
ASII GIII PIIE I	AAA TTA ACT AG Lys Leu Thr Se 900	905	Asn Ser Lys	Ile Arg Val 910	Thr
915	AAT ATC ATA TT Asn Ile Ile Ph	920	/al Phe Leu	Asp Phe Ser 925	Val
930	ATA AGA ATA CC Ile Arg Ile Pro 93	D Lys Tyr L	ys Asn Asp 940	Gly Ile Gln	Asn
945	AT GAA TAT AC Asn Glu Tyr Th: 950	: Ile Ile A	sn Cys Met 955	Lys Asn Asn	Ser 960
GIY ITP LYS I	TA TCT ATT AGG Ele Ser Ile Arg 965	J Gly Asn A 9	rg Ile Ile 70	Trp Thr Leu 975	Ile
Asp lie Asn G 9	GA AAA ACC AAA lly Lys Thr Lys 180	Ser Val P 985	he Phe Glu	Tyr Asn Ile 990	Arg
Glu Asp lie S 995	CA GAG TAT ATA er Glu Tyr Ile	Asn Arg T:	rp Phe Phe	Val Thr Ile 1005	Thr
Asn Asn Leu A 1010	AT AAC GCT AAA sn Asn Ala Lys 101	Ile Tyr I	le Asn Gly : 1020	Lys Leu Glu	Ser
Asn Thr Asp I 1025	TT AAA GAT ATA le Lys Asp Ile 1030	Arg Glu Va	al Ile Ala i 1035	Asn Gly Glu	Ile 1040
ATA TTT AAA T Ile Phe Lys L	TA GAT GGT GAT eu Asp Gly Asp 1045	Ile Asp A	GA ACA CAA 1 rg Thr Gln 1 050	TTT ATT TGG Phe Ile Trp 1055	ATG 3168 Met
Lys Tyr Phe S	GT ATT TTT AAT er Ile Phe Asn 060	ACG GAA TT Thr Glu Le 1065	TA AGT CAA 1 eu Ser Gln 9	TCA AAT ATT Ser Asn Ile 1070	GAA 3216 Glu
GAA AGA TAT A Glu Arg Tyr L 1075	AA ATT CAA TCA ys Ile Gln Ser	TAT AGC GA Tyr Ser Gl 1080	lu Tyr Leu I	AAA GAT TTT ' Lys Asp Phe ' 1085	TGG 3264 Trp

GGA Gly	AAT Asn 109	Pro	TTA Leu	ATG Met	TAC	AAT Asn 109	Lys	GAA Glu	TAT	TAT Tyr	ATG Met 110	TTT Phe 0	AAT Asn	GCG Ala	GGG Gly	3	312
AAT Asn 110	Lys	AAT Asn	TCA Ser	TAT Tyr	ATT Ile 1110	Lys	CTA Leu	AAG Lys	AAA Lys	GAT Asp 111	Ser	CCT Pro	GTA Val	GGT Gly	GAA Glu 1120	3	360
ATT Ile	TTA Leu	ACA Thr	CGT Arg	AGC Ser 1125	Lys	TAT Tyr	AAT Asn	CAA Gln	AAT Asn 1130	Ser	AAA Lys	TAT Tyr	ATA Ile	AAT Asn 1139	Tyr	3	408
AGA Arg	GAT Asp	·TTA Leu	TAT Tyr 1140	Ile	GGA Gly	GAA Glu	AAA Lys	TTT Phe 1145	Ile	ATA Ile	AGA Arg	AGA Arg	AAG Lys 1150	Ser	AAT Asn	3	456
TCT Ser	CAA Gln	TCT Ser 1155	Ile	AAT Asn	GAT Asp	GAT Asp	ATA Ile 1160	Val	AGA Arg	AAA Lys	GAA Glu	GAT Asp 1165	Tyr	ATA Ile	TAT Tyr	3 9	504
CTA Leu	GA															35	509

(2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1169 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe
85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly
165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 200 Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 265 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 330 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 345 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 395 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 455 Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys

Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 540

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 545 550 555 560

Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 570 575

Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 580 585 590

Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 595 600 605

Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 620

Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 635 630 635

Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655

Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665 670

Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685

Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 695 700

Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720

Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr
725 730 735

Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740 745 750

Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755 760 765

Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 775 780

Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 785 790 795 800

Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 810 815

Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 820 825 830

Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 835 840 845

Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile 850 855 860

Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly 865 870 875 880 Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 885 890 895

Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr 900 905 910

Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 915 920 925

Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 930 935 940

Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 945 950 955 960

Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile 965 970 975

Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 980 985 990

Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 995 1000 1005

Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1010 1015 1020

Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 1040

Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met 1045 1050 1055

Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1060 1065 1070

Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080 1085

Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly 1090 1095 1100

Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu 1105 1110 1115 1120

Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr 1125 1130 1135

Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn 1140 1145 1150

Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr 1155 1160 1165

Leu

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2574

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

										-								
AT Me	G CC t Pr 1	CA G	TT A	CA hr	ATA Ile	a As	T AA n As	T TI n Ph	T AA le As	n Ty	T AA T As O	T GA n As	AT Co	CT A	TT (GA1 Asp 15	AAT Asn	48
AS	n As	11 1.	re i	20	Met	. mei	c GT	u Pr	2	o Ph 5	e Ala	a Ar	g G	Ly Ti	hr (Gly	AGA Arg	96
ry.	L IY	3	5 A	14	Pne	Lys	5 110	9 Th 4	r As	p Ar	T AT:	e Tr	p Il 4	e I] 5	Le E	Pro	Glu	144
vr.	5 TY	0	IF P	ne	GIÀ	ıyı	55 55	s Pro	5 G11	u Asp	r TTI Phe	As: 6€	n Ly O	s Se	er S	er	Gly	192
65	5	e AS	n A	cg .	ASP	70	Cys	i GI	тул	с Туг	GAT Asp 75) Pro	o As	р Ту	r L	eu	Asn 80	240
1111	ASI	1 AS	ը - /	75 1	85	Asn	11e	Phe	Leu	90		Met	: Il	e Ly	s L	eu 95	Phe	288
ASII	Arc	; L1:	10	0	ser	Lys	Pro	Leu	105	Glu	AAG Lys	Leu	ı Leı	1 Gl	u Mo O	et	Ile	336
116	ASI	115	2 11	e £	ro	Tyr	Leu	120	Asp	Arg	CGT	Val	125	Le	u G	Lu	Glu	384
rne	130	INI	. AS	n 1	ııe	Ala	Ser 135	Vaļ	Thr	Val	AAT Asn	Lys 140	Leu	ı Ile	e Se	er.	Asn	432
145	GIY	GIU	va.	1 G	ilu .	Arg 150	Lys	Lys	Gly	Ile	TTC Phe 155	Ala	Asn	Let	ıIl	.e	lle 160	480
Pne	GIY	PIC	GI	1	65	Val	Leu	Asn	Glu	170	GAG Glu	Thr	Ile	qeA	11 17	e (Gly	528
116	GIN	Asn	180	5 P)	he A	Ala	Ser	Arg	Glu 185	Gly	TTC Phe	Gly	Gly	Ile 190	Me	t (Gln	576
Met	ьys	195	Cys	s Pi	ro (31u	Tyr	Val 200	Ser	Val	TTT Phe	Asn	Asn 205	Val	G1:	n G	lu	624
Asn	Lys 210	GIÀ	Ala	. S€	er 1	lle	Phe 215	Asn	Arg	Arg		Tyr 220	Phe	Ser	Ası	9 9	ro	672
GCC Ala 225	TTG Leu	ATA Ile	TTA	AT Me	et H	AT (lis (GAA Glu	CTT Leu	ATA Ile	His	GTT ' Val : 235	TTA Leu	CAT His	GGA Gly	TT/ Let	1 T	AT yr 40	720

GGC Gly	ATT	AAA Lys	GTA Val	GAT Asp 245) Asp	TTA Leu	CCA Pro	ATT	GTA Val 250	Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT	768
TTT Phe	Met	CAA Gln	TCT Ser 260	Inr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	Thr	TTT	816
GGA Gly	GGA Gly	CAA Gln 275	. Asp	CCC	AGC Ser	ATC Ile	ATA Ile 280	Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
TAT Tyr	GAT Asp 290	ьys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT A sp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
TAA neA	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Gl <u>u</u>	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536

GAT Asp	TTT Phe	AAT Asr 515	ı Val	A GAT L Asp	GTT Val	CCA Pro	GT/ Val	L Ty	r GAZ	AAA Lys	CAV	A CCC n Pro 525	Al.	T AT a Il	A AAA e Lys		1584
AAA Lys	ATT	Phe	ACA Thr	A GAT	GAA	AAT Asn 535	Thr	C ATO	TTI Phe	CAF Glr	TAT Ty: 540	Let	A TAC	TC r Se	T CAG r Gln	,	1632
ACA Thr 545	Phe	CCI Pro	CTA Leu	GAT Asp	ATA	Arg	GAT Asp	T ATA	AGT Ser	TTA Leu 555	Thr	TCI Ser	TC/	TT Ph	GAT Asp 560		1680
GAT Asp	GCA Ala	TTA Leu	TTA Leu	TTT Phe 565	Ser	AAC Asn	AAA Lys	GT1	TAT Tyr 570	Ser	TTI Phe	TT1	TCT Ser	T ATC Met 575	G GAT Asp		1728
TAT Tyr	ATT Ile	AAA Lys	ACT Thr 580	Ala	AAT Asn	AAA Lys	GTG Val	GTA Val 585	Glu	GCA Ala	GGA Gly	TTA Leu	TTI Phe 590	Ala	GGT Gly		17 7 6
TGG Trp	GTG Val	AAA Lys 595	CAG Gln	ATA Ile	GTA Val	AAT Asn	GAT Asp 600	Phe	GTA Val	ATC Ile	GAA Glu	GCT Ala 605	AAT Asn	AAA Lys	AGC Ser		1824
AAT Asn	ACT Thr 610	ATG Met	GAT Asp	AAA Lys	ATT Ile	GCA Ala 615	GAT Asp	ATA Ile	TCT Ser	CTA Leu	ATT Ile 620	GTT Val	CCT Pro	TAT	ATA Ile		1872
GGA Gly 625	TTA Leu	GCT Ala	TTA Leu	AAT Asn	GTA Val 630	GGA Gly	AAT Asn	GAA Glu	ACA Thr	GCT Ala 635	AAA Lys	GGA Gly	AAT Asn	TTI Phe	GAA Glu 640		1920
AAT Asn	GCT Ala	TTT Phe	GAG Glu	ATT Ile 645	GCA Ala	GGA Gly	GCC Ala	AGT Ser	ATT Ile 650	CTA Leu	CTA Leu	GAA Glu	TTT Phe	ATA Ile 655	CCA Pro		1968
GAA Glu	CTT Leu	TTA Leu	ATA Ile 660	CCT Pro	GTA Val	GTT Val	GGA Gly	GCC Ala 665	TTT Phe	TTA Leu	TTA Leu	GAA Glu	TCA Ser 670	TAT	ATT Ile		2016
GAC Asp	AAT Asn	AAA Lys 675	AAT Asn	AAA Lys	ATT Ile	ATT Ile	AAA Lys 680	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 685	TTA Leu	ACT Thr	AAA Lys		2064
AGA Arg	AAT Asn 690	Glu	Lys	Trp	Ser	GAT Asp 695	Met	Tyr	Gly	Leu	Ile	Val	GCG Ala	CAA Gln	TGG Trp		2112
CTC Leu 705	TCA Ser	ACA Thr	GTT Val	AAT Asn	ACT Thr 710	CAA Gln	TTT Phe	TAT Tyr	ACA Thr	ATA Ile 715	AAA Lys	GAG Glu	GGA Gly	ATG Met	TAT Tyr 720		2160
AAG Lys	GCT Ala	TTA Leu	AAT Asn	TAT Tyr 725	CAA Gln	GCA Ala	CAA Gln	GCA Ala	TTG Leu 730	GAA Glu	GAA Glu	ATA Ile	ATA Ile	AAA Lys 735	TAC Tyr		2208
AGA Arg	TAT Tyr	AAT Asn	ATA Ile 740	TAT Tyr	TCT Ser	GAA Glu	AAA Lys	GAA Glu 745	AAG Lys	TCA Ser	AAT Asn	ATT Ile	AAC Asn 750	ATC Ile	GAT Asp		2256
TTT Phe	Asn	GAT Asp 755	ATA Ile	AAT Asn	TCT Ser	Lys	CTT Leu 760	AAT Asn	GAG Glu	GGT Gly	ATT Ile	AAC Asn 765	CAA Gln	GCT Ala	ATA Ile		2304
GAT Asp	AAT Asn 770	ATA Ile	AAT Asn	AAT Asn	Phe	ATA . Ile . 775	AAT Asn	GGA Gly	TGT Cys	Ser	GTA Val 780	TCA Ser	TAT Tyr	TTA Leu	ATG Met		2352

_	AA ys 85	AAA Lys	ATG Met	ATT Ile	CCA Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400
I	CT hr	CTC	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA Leu	AAT Asn	TAT Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448
T L	TG eu	ATT Ile	GGA Gly	AGT Ser 820	GCA Ala	GAA Glu	TAT Tyr	GAA Glu	AAA Lys 825	TCA Ser	AAA Lys	GTA Val	AAT Asn	AAA Lys 830	TAC Tyr	TTG Leu	2496
A L	AA ys	1111	ATT Ile 835	ATG Met	CCG Pro	TTT Phe	GAT Asp	CTT Leu 840	TCA Ser	ATA Ile	TAT Tyr	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	2544
C.	eu	ATA Ile 850	GAA Glu	ATG Met	TTT Phe	Asn	AAA Lys 855	TAT Tyr	AAT Asn	AGC Ser							2574

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 858 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

 (xi)
 SEQUENCE DESCRIPTION:
 SEQ ID No:
 22:

 Met 1
 Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 15

 Asn Asn Ile Ile Met Met Glu Pro Pro Pro Pro Phe Ala Arg Gly Thr Gly Arg 20

 Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu Asp Phe Asn Lys Ser Ser Gly 55

 Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 55

 Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 80

 Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 95

 Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 110

 Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu Glu 115

 Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130

 Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
180 185 190

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 215 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 250 Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 330 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 345 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 360 Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 375 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 390 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 440 Val Asp`Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 505 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 520 Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln

535

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 550 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser 855

- (2) INFORMATION FOR SEQ ID NO: 23:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1644 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..1644

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

ATO Met	Pro	GTI Val	ACA Thr	ATA Ile	Asn	AAT Asn	TTI Phe	AA1 Asr	TAI Tyr 10	Asr	GAT Asp	CCT Pro	ATI	GAT Asp 19	AAT Asn		48
AAT Asn	AAT Asn	ATT	ATT Ile 20	Met	ATG Met	GAG Glu	CCT	CCA Pro 25	Phe	GCG Ala	AGA Arg	GGT Gly	ACG Thr	Gly	AGA Arg		96
TAT Tyr	TAT Tyr	AAA Lys 35	Ala	TTT Phe	AAA Lys	ATC Ile	ACA Thr 40	Asp	CGT Arg	ATT Ile	TGG	ATA Ile 45	ATA Ile	CCG Pro	GAA Glu	1	144
AGA Arg	TAT Tyr 50	Thr	TTT Phe	GGA Gly	TAT Tyr	AAA Lys 55	CCT Pro	GAG Glu	GAT Asp	TTT Phe	AAT Asn 60	Lys	AGT Ser	TCC Ser	GGT Gly	1	L92
ATT Ile 65	TTT Phe	AAT Asn	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT Tyr	TAT Tyr	GAT Asp 75	CCA Pro	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 80	2	240
ACT Thr	AAT Asn	GAT Asp	AAA Lys	AAG Lys 85	AAT Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC Ile	AAG Lys	TTA Leu 95	TTT Phe	2	88
AAT Asn	AGA Arg	ATC Ile	AAA Lys 100	TCA Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT Ile	3	36
ATA Ile	AAT Asn	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu	3	84
TTT Phe	AAC Asn 130	ACA Thr	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn	4	32
CCA Pro 145	GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	TAA Asn	TTA Leu	ATA Ile	ATA Ile 160	4	80
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly	52	28
ATA Ile	CAA Gln	Asn	CAT His 180	Phe	Ala	Ser	Arg	Glu	Gly	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln	51	76
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	62	24
AAC Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT Arg	GGA Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro	67	72

GCC Ala 225	Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAA Glu	CTT Leu	ATA Ile	CAT His	GTT Val 235	Leu	CAT His	GGA Gly	TTA Leu	TAT Tyr 240	720
GGC Gly	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA Leu	CCA Pro	ATT	GTA Val 250	Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe	. 768
TTT Phe	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	ACA Thr	TTT Phe	816
GGA Gly	GGA Gly	CAÁ Gln 275	GAT Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
ryr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488

- 103 -

TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	,	1536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys		1584
AAA Lys	ATT Ile 530	TTT	ACA Thr	GAT Asp	Gru	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln		1632
ACA Thr 545	TTT Phe															:	1644

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 548 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg
20 25 30

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe
85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
180 185 190

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205

Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 215 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 235 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 265 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 360 Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 390 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 410 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 440 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 490 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 505 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln Thr Phe Pro Leu 545

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2616 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AT(G CA C G1:	G TT n Ph	C GT e Va	G AA l As	C AAG n Lys	G CAC	TTO Phe	C AAC Asi	TA: 1 Ty:	r Lys	G GA	C CC P Pr	T GT.	A AA l As l	C GGT n Gly 5	4.6
GT: Val	GA L As	C AT p Il	T GC e Al 2	u .y.	C ATO	AAA Lys	ATT Ile	CCA Pro	AST	GCC Ala	GG(CAC Gl:	G ATO	G1:	G CCG n Pro	96
GTC Val	Lys	G GC' S Ala	~ ~ 11.	C AAG e Lys	G ATT	CAT His	AAC Asn 40	Lys	ATC	TGG	GT1 Val	ATT 116	Pro	GA Glu	A CGC	144
GAT Asp	ACA Thr		Γ ACC ⊇ Thi	G AAC C Asr	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	Pro	CCG Pro	CCC Pro	G GAA Glu	192
65	2,3	GII	. vai	. Plu	70	ser	lyr	Tyr	· Asp	Ser 75	Thr	Tyr	Leu	Ser	ACA Thr 80	240
пор	71311	Giu	. шуз	85	ASII	TAC Tyr	Leu	гÀг	90	Val	Thr	Lys	Leu	Phe 95	Glu	288
		-7-	100	****	vəħ	CTG Leu	GIY	105	мес	Leu	Leu	Thr	Ser 110	Ile	Val	336
9	O. J	115	110	rne	ΙΙĐ	GGT Gly	120	ser	Tnr	He	Asp	Thr 125	Glu	Leu	Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA . Val	ATC Ile	ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
-10	J 111	rne	GIU	165	гуз	AGC Ser	rne (GIY .	H15 (Glu '	Val	Leu	Asn	Leu 175	Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT . Ser	ACT (Thr (STIL.	TAC A Tyr :	ATT (CGT :	TTC . Phe	Ser	CCA Pro 190	GAC Asp	TTC Phe	576

ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
Tyr	Leu	Leu	Ser	GAA Glu 325	Asp	Thr	Ser	Gly	Lys 330	Phe	Ser	Val	Asp	Lys 335	Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392

46	5		.	Lu A	4°	70	ie 11	II AS	on As	эр Le . 47	'5	AT AA sn Ly	s G]	ly G	lu	Glu 480	1440
ATT Ile	T AC	A TO	T GA	AT AC sp Th 48	II AS	AT AT	A GA e Gl	A GO .u Al	A GC a Al 49	a GI	A GA u Gl	A AA Lu As	T AT	e S	GT er 95	TTA Leu	1488
GAT Asp	TTZ Lei	A AT	A CA e Gl 50	n Gi	A TA n Ty	T TA	T TI r Le	A AC u Th 50	r Ph	T AA	T TI n Ph	T GA	T AA p As 51	n G	AA lu	CCT Pro	1536
GAA Glu	AA? Asi	T AT 1 Il 51	e 3e	A AT	A GA e Gl	A AA' u Asi	T CT n Le 52	u se	A AG r Se	T GAG	C AT	T ATE	e Gl	C CA Y GI	AA In	TTA Leu	1584
GAA Glu	Leu 530	i rie	G CC t Pr	T AA o As	T AT	A GAZ e Gli 539	1 Ar	A TT g Ph	T CC	T AAT O Asr	GG. 1 Gl: 54	A AA y Lys 0	A AAG	G TA s Ty	T (GAG Glu	1632
TTA Leu 545	GAT Asp	Ly:	A TA	T AC	T ATO	c Pue	CAT His	T TA	r CT	CGT Arg 555	Ala	T CAA a Gln	GA/	A TT	e (GAA Glu 560	1680
CAT His	GGT Gly	AA/ Lys	A TC	AG0 Ar9 569	3 776	r gcī ⊇ Ala	Let	A ACA	A AAT Asr 570	ı Ser	GTT Val	CAA 1 L Asn	GAA Glu	GC Al 57	a I	TA Leu	, 1728
TTA Leu	AAT Asn	CCT	AG1 Se1 580	. wri	r GTT g Val	TAT Tyr	ACA Thr	TT1 Phe 585	Phe	TCT Ser	TCA Ser	A GAC	TAT Tyr 590	Va.	A A l L	AG ys	1776
AAA Lys	GTT Val	AAT Asn 595	. цув	GCI Ala	ACG Thr	GAG Glu	GCA Ala 600	Ala	'ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GT/ Val	A G	AA lu	1824
GIII	TTA Leu 610	GTÁ Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAÁ Glu 620	GTA Val	AGT Ser	ACT Thr	A	CG hr	1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	A	CT la 10	1920
TTA . Leu .	AAT Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	Asp	Asp	Phe	GTA Val	Gly	Ala	T: Le	ra eu	1968
ATA :	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT Ile	GC Al	A .a	2016
ATA (-10	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AA Ly	.G 's	2064
GTT (Val I	TA Leu 190	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	Leu :	AGT Ser 700	AAA . Lys .	AGA Arg	AAT Asn	GA G1	A u	2112
AAA T Lys T 705	GG (GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT Tyr	ATA Ile	GTA Val	ACA I Thr I 715	AAT Asn	TGG '	TTA : Leu .	GCA Ala	AA Ly 72	s	2160
GTT A	AT 1	ACA Thr	CAG Gln	ATT Ile 725	GAT Asp	CTA /	ATA Ile	Arg	AAA Lys 730	AAA A Lys N	ATG Met	AAA (Lys (Glu ,	GCT Ala 735	TT. Le	A u	2208

GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp	2304
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA Ile	2352
AAT Asn 785	AAA Lys	TTT- Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2400
ATC Ile	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
						ATA Ile										2496
						GAT Asp										2544
						AAA Lys 855										2592
		ACT Thr				AAG Lys	TAA *									2616

- (2) INFORMATION FOR SEQ ID NO: 26:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 872 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 135 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 250 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 345 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 360 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 440 Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 535 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 570 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 585 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 600 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 615 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 665 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 760 Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 825

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser

Thr Phe Thr Glu Tyr Ile Lys . 870

- (2) INFORMATION FOR SEQ ID NO: 27:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs

 - (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

ATGCCGGTTA CCATCAACAA CTTCAACTAC AACGACCCG	A TCGACAACA	A CAACATCATC	60
ATGATGGAAC CGCCGTTCGC ACGTGGTACC GGTCGTTAC	T ACAAGGCTT	r caagatcacc	120
GACCGTATCT GGATCATCCC GGAACGTTAC ACCTTCGGT	T ACAAACCTG	A GGACTTCAAC	180
AAGAGTAGCG GGATTTTCAA TCGTGACGTC TGCGAGTAC	T ATGATCCAG	A TTATCTGAAT	240
ACCAACGATA AGAAGAACAT ATTCCTTCAG ACTATGATCA	AGTTATTA	TAGAATCAAA	300
TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA	ATGGTATACO	TTATCTTGGA	360
GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG	CTAGTGTAAC	TGTTAATAAA	420
TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA	TTTTCGCAAA	ATAATAATA	480
TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG	ATATAGGTAT	ACAAAATCAT	540
TTTGCATCAA GGGAAGGCTT CGGGGGTATA ATGCAAATGA	AGTTTTGCCC	AGAATATGTA	600
AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA	TATTTAATAG	ACGTGGATAT	660
TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTTATAC	ATGTTTTACA	TGGATTATAT	720
GGCATTAAAG TAGATGATTT ACCAATTGTA CCAAATGAAA	AAAAATTTTT	TATGCAATCT	780
ACAGATGCTA TACAGGCAGA AGAACTATAT ACATTTGGAG	GACAAGATCC	CAGCATCATA	840
ACTCCTTCTA CGGATAAAAG TATCTATGAT AAAGTTTTGC	AAAATTTTAG	AGGGATAGTT	900
GATAGACTTA ACAAGGTTTT AGTTTGCATA TCAGATCCTA	ACATTAATAT	TAATATATAT	960
AAAAATAAAT TTAAAGATAA ATATAAATTC GTTGAAGATT	CTGAGGGAAA	ATATAGTATA	1020
GATGTAGAAA GTTTTGATAA ATTATAAAA AGCTTAATGT	TTGGTTTTAC	AGAAACTAAT	1080
ATAGCAGAAA ATTATAAAAT AAAAACTAGA GCTTCTTATT	TTAGTGATTC	CTTACCACCA	1140
GTAAAAATAA AAAATTTATT AGATAATGAA ATCTATACTA			1200

TCTGATAAAG	ATATGGAAAA	AGAATATAGA	GGTCAGAATA	AAGCTATAAA	TAAACAAGCT	1260
TATGAAGAAA	TTAGCAAGGA	GCATTTGGCT	GTATATAAGA	TACAAATGTG	TAAAAGTGTT	1320
AAAGCTCCAG	GAATATGTAT	TGATGTTGAT	AATGAAGATT	TGTTCTTTAT	AGCTGATAAA	1380
AATAGTTTTT	CAGATGATTT	ATCTAAAAAC	GAAAGAATAG	AATATAATAC	ACAGAGTAAT	1440
TATATAGAAA	ATGACTTCCC	TATAAATGAA	TTAATTTTAG	ATACTGATTT	AATAAGTAAA	1500
ATAGAATTAC	CAAGTGAAAA	TACAGAATCA	CTTACTGATT	TTAATGTAGA	TGTTCCAGTA	1560
TATGAAAAAC	AACCCGCTAT	AAAAAAAATT	TTTACAGATG	AAAATACCAT	CTTTCAATAT	1620
TTATACTCTC	AGACATTTCC	TCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
GATGCATTAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTT	CTATGGATTA	TATTAAAACT	1740
GCTAATAAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTTGTAATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
GTTCCTTATA	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
AATGCTTTTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CCTGTAGTTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	АТААААТАА	AATTATTAAA	2040
ACAATAGATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
GTAGCGCAAT	GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AAGGCTTTAA	ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	ATATAATATA	2220
TATTCTGAAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280
aatgagggta	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
TCATATTTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA	AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
TCAATATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2574 base pairs
 (B) TYPE: nucleic acid

 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

60	TAATATTATT	TTGATAATAA	AATGATCCTA	TTTTAATTAT	CAATAAATAA	ATGCCAGTTA
120	TAAAATCACA	ATAAAGCTTT	GGGAGATATT	GAGAGGTACG	CTCCATTTGC	ATGATGGAGC
180	GGATTTTAAT	ATAAACCTGA	ACTITIGGAT	GGAAAGATAT	GGATAATACC	GATCGTATTT
240	TTACTTAAAT	ATGATCCAGA	TGTGAATATT	TAGAGATGTT	GTATTTTTAA	AAAAGTTCCG

ACTAATGATA AAAAGAATAT ATTTTTACAA ACAATGATCA AGTTATTTAA TAGAATCAAA	300
TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA	360
GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA	420
TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAAA TTTAATAATA	480
TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAAATCAT	540
TTTGCATCAA GGGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTTGCCC AGAATATGTA	600
AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACGTGGATAT	660
TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTCATCC ACGTCCTCCA CGGTCTCTAC	720
GGTATCAAAG TAGACGACCT CCCGATCGTC CCGAACGAAA AAAAATTCTT CATGCAGAGC	780
ACCGACGCAA TCCAGGCAGA AGAACTCTAC ACCTTCGGTG GTCAGGACCC GAGCATCATC	840
ACCCCGAGCA CCGACAAAAG CATCTACGAC AAAGTCCTCC AGAACTTCCG TGGTATCGTC	900
GACCGTCTCA ACAAAGTCCT CGTCTGCATC AGCGACCCGA ACATCAACAT CAACATCTAC	960
AAAAACAAAT TCAAAGACAA ATACAAATTC GTCGAAGACA GCGAAGGTAA ATACAGCATC	1020
GACGTCGAGA GCTTCGACAA ACTCTACAAA AGCCTCATGT TCGGTTTCAC CGAAACCAAC	1080
ATCGCAGAAA ACTACAAAAT CAAAACCCGT GCAAGCTACT TCAGCGACAG CCTCCCGCCG	1140
GTCAAAATCA AAAACCTCCT CGACAACGAA ATCTACACCA TCGAAGAAGG TTTCAACATC	1200
AGCGACAAAG ACATGGAAAA AGAATACCGT GGTCAGAACA AAGCAATCAA CAAACAAGCT	1260
TACGAAGAAA TCAGCAAAGA ACACCTCGCA GTCTACAAAA TCCAGATGTG CAAAAGCGTC	1320
AAAGCACCGG GTATCTGCAT CGACGTTGAC AACGAAGACC TCTTCTTCAT CGCAGACAAA	1380
AACAGCTTCA GCGACGACCT CAGCAAAAAC GAACGTATCG AATACAACAC CCAGAGCAAC	1440
TACATCGAAA ACGACTTCCC GATCAACGAA CTCATCCTCG ACACCGACCT CATCAGCAAA	1500
ATCGAACTCC CGAGCGAAAA CACCGAAAGC CTCACCGACT TCAACGTTGA CGTCCCGGTC	1560
TACGAAAAAC AGCCGGCAAT CAAAAAAATC TTCACCGACG AAAACACCAT CTTCCAGTAC	1620
CTCTACAGCC AGACCTTCCC GCTAGATATA AGAGATATAA GTTTAACATC TTCATTTGAT	1680
GATGCATTAT TATTTTCTAA CAAAGTTTAT TCATTTTTTT CTATGGATTA TATTAAAACT	1740
GCTAATAAAG TGGTAGAAGC AGGATTATTT GCAGGTTGGG TGAAACAGAT AGTAAATGAT	1800
TTTGTAATCG AAGCTAATAA AAGCAATACT ATGGATAAAA TTGCAGATAT ATCTCTAATT	1860
GTTCCTTATA TAGGATTAGC TTTAAATGTA GGAAATGAAA CAGCTAAAGG AAATTTTGAA	1920
AATGCTTTTG AGATTGCAGG AGCCAGTATT CTACTAGAAT TTATACCAGA ACTTTTAATA	1980
CCTGTAGTTG GAGCCTTTTT ATTAGAATCA TATATTGACA ATAAAAATAA AATTATTAAA	2040
ACAATAGATA ATGCTTTAAC TAAAAGAAAT GAAAAATGGA GTGATATGTA CGGATTAATA	2100
GTAGCGCAAT GGCTCTCAAC AGTTAATACT CAATTTTATA CAATAAAAGA GGGAATGTAT	2160
AAGGCTTTAA ATTATCAAGC ACAAGCATTG GAAGAAATAA TAAAATACAG ATATAATATA	2220
TATTCTGAAA AAGAAAAGTC AAATATTAAC ATCGATTTTA ATGATATAAA TTCTAAACTT	2280

- 114 -

AATGAGGGTA	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
TCATATTTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA	AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
CAATATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

CLAIMS

- 1. A polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and wherein said second domain is adapted (i) to translocate the polypeptide into a cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into a cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of clostridial neurotoxin precursor that can be converted into toxin by proteolytic action.
- 2. A polypeptide according to Claim 1 wherein said first domain comprises a clostridial toxin light chain.
- 3. A polypeptide according to Claim 1 wherein said first domain comprises a fragment or variant of a clostridial toxin light chain.
- 4. A polypeptide according to Claim 2 or 3 wherein the clostridial toxin is a botulinum toxin.
- 5. A polypeptide according to any preceding claim wherein the first domain exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
- 6. A polypeptide according to any preceding claim wherein said second domain comprises a clostridial toxin heavy chain H_N portion.
- 7. A polypeptide according to any of Claims 1-5 wherein said second domain comprises a fragment or variant of a clostridial toxin heavy chain H_N portion.
- 8. A polypeptide according to Claim 6 or 7 wherein the clostridial toxin is a

botulinum toxin.

- 9. A polypeptide according to any of Claims 1-8 further comprising a third domain adapted for binding of the polypeptide to a cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that bind to a cell.
- 10. A polypeptide according to Claim 9 wherein said third domain is for binding the polypeptide to an immunoglobulin.
- 11. A polypeptide according to Claim 10 wherein said third domain is a tandem repeat synthetic IgG binding domain derived from domain β of Staphylococcal protein A.
- 12. A polypeptide according to Claim 9 wherein said third domain comprises an amino acid sequence that binds to a cell surface receptor.
- 13. A polypeptide according to Claim 12 wherein said third domain is insulin-like growth factor-1 (IGF-1).
- 14. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and a portion designated H_{N} of a botulinum toxin heavy chain.
- 15. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A.
- 16. A polypeptide according to Claim 15 wherein the botulinum toxin type A light chain variant has at residue 2 a glutamate, at residue 26 a lysine and at residue 27 a tyrosine.

- 17. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
- 18. A polypeptide according to any of Claims 1-13 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.
- 19. A polypeptide according to Claim 18 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 20. A polypeptide according to Claim 15 or 16 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
- 21. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 22. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 23. A polypeptide according to Claim 17 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
- 24. A polypeptide according to Claim 23 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 25. A polypeptide according to any of Claims 14-24 lacking a portion designated

H_c of a botulinum toxin heavy chain.

- 26. A polypeptide comprising a botulinum toxin light chain and a fragment of a botulinum toxin heavy chain, said fragment being not capable of binding to cell surface receptors.
- 27. A polypeptide according to Claim 26 lacking an intact portion designated $H_{\rm c}$ of a botulinum toxin heavy chain.
- 28. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
- 29. A polypeptide according to Claim 28 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
- 30. A polypeptide according to Claim 28 or 29 comprising a variant of a clostridial toxin heavy chain H_N portion and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain H_N portion.
- 31. A polypeptide according to Claim 28, 29 or 30 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotides coding for the cleavage site.
- 32. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-31 with (b) a second polypeptide being a polypeptide or oligopeptide adapted for binding to an affinity matrix so as to enable purification of the fusion protein using said matrix.
- 33. A fusion protein according to Claim 32 wherein said second polypeptide is

adapted to bind to a chromatography column, such as an affinity matrix of glutathione Sepharose.

- 34. A fusion protein according to Claim 32 or 33 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.
- 35. A composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the botulinum toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*.
- 36. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a positive control in a toxin assay.
- 37. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a vaccine against clostridial toxin.
- 38. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for *in vivo* use.
- 39. A pharmaceutical composition comprising a composition according to Claim 35, a polypeptide according to any of claims 1-31 or a fusion protein according to Claim 32, 33 or 34, in combination with a pharmaceutically acceptable carrier.
- 40. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-34.
- 41. A nucleic acid encoding a polypeptide or a fusion protein according to Claim

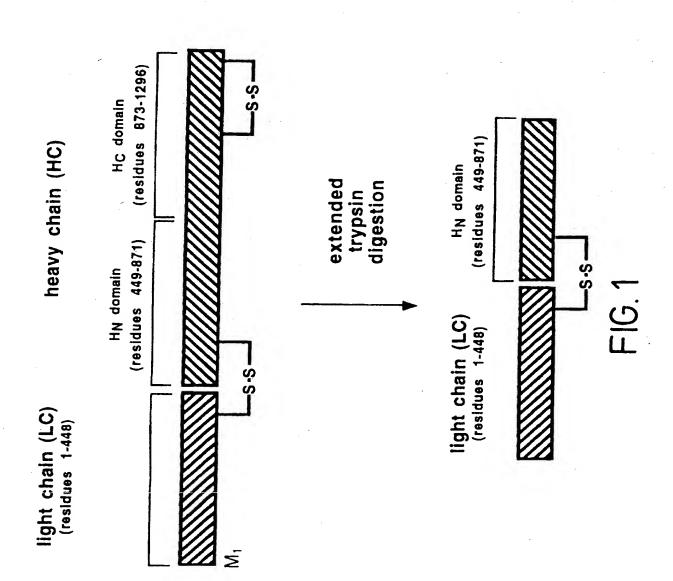
40 and comprising nucleotides encoding residues 1-448 of a botulinum toxin type A light chain.

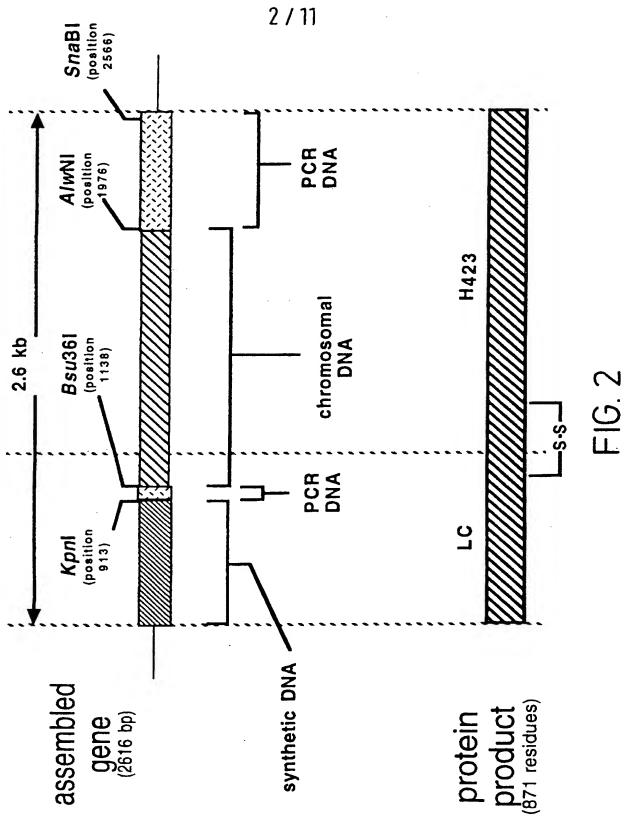
- 42. A nucleic acid according to Claim 40 or 41 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain H_N domain.
- 43. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 and comprising nucleotides encoding residues 1-470 of a botulinum toxin type B light chain.
- 44. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 or 43 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain H_{N} domain.
- 45. A nucleic acid according to any of Claims 40-44 comprising nucleotides encoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.
- 46. A nucleotide according to Claim 45 obtainable by modification of a nucleotide encoding a polypeptide or fusion protein according to any of claims 1-34 so as to introduce said cleavage site.
- 47. A DNA according to any of claims 40-46.
- 48. A DNA selected from SEQ ID No:s 1, 8, 10, 12, 14, 16, 18, 23 and 24.
- 49. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid according to any of Claims 40-48 and recovering the polypeptide.
- 50. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid encoding a fusion protein

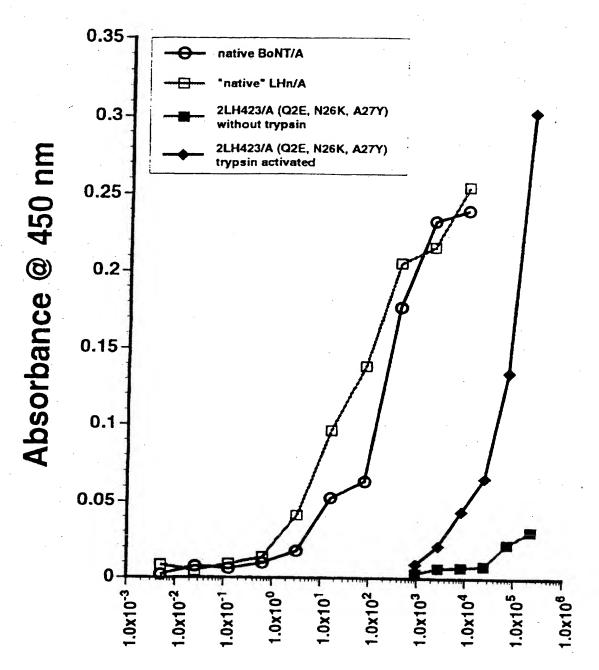
according to Claim 32, 33 or 34, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.

- 51. A method of manufacture according to Claims 49 or 50 in which the nucleic acid is DNA.
- 52. A cell expressing a polypeptide or fusion protein according to any of Claims 1-34.









Protein concentration (ng/ml)

FIG. 3

			4/1	7		
2		₹	<u> </u>			
DIAYIKIPNAGQMQPV (Seq I.D. 2)		K Y G Q M Q P V (Seq I.D. 4)	Q P V (Seq I.D. 6)			
_		<u> </u>				
Se		Se	Se			
<u> </u>		<u> </u>		>	_	
0		0	_	۵	۵	
~		~	8	0	~	
Σ		Σ	Σ	Σ	Σ	
0		0	0	∑	0	
S		5	¥	S	N A G Q M Q	
∢		>	>	Ø	4	
Z		×	×	Z	Z	
۵.		A I P	٥	۵	۵	
H		H	I K I	H	YIKI	
×		×	×	×	×	
H		H		н	H	
>		>	>	>-	>-	
∢		∢	∢	∢	⋖	
—	ا 9	I (I Q	I Q	I O	
>	ш	<u> </u>	>	>	>	
5	11.	6			9	
z	ш Ш	z	Z	⊍ Z	z	
>		>		>	>	
م		۵.	0 V	۵	· d	
<u> </u>	¥	۵	۵	Q	0	
×	0	×	×	¥	2	
>	TRLQKLL	>	>	F N Y K D	>-	
z	œ		Z	Z	Z	
ட	-	Z L	ட	u.	44.	
0	S	0	0	0	0	
¥	F	×	×	¥	×	
Z	X	Z	Z	Z	Z	
>	I	>	>	>	>	
M Q F V N K Q	GSPGIHMT 1	G T M E F V N K Q	G S M E F V N K 123) H C C C C C C C C C C C C C C C C C C	M F V N K Q	
<u>~</u>]	6	ш	ш	~		
≆ ∺	۵.	T M 24	G S M 123	≖ ∺	∑ ⊣	
	S	-	2 2	•		
	6-		0 4			
				Q		
		Έ	ξ.	39 199	~	
		A ₂	2LH423/A (Q2E, N26K, A27Y)	Native BoNT/A, C. botulinum 2169 Thompson <i>et al.</i> 1990	Native BoNT/A, C. botulinum 62A Binz et al.1990	
∢	a	ᄎ	Ž	Native BoNT/A, C. botulinum 21 Thompson <i>et al</i>	Native BoNT/A, C. botulinum 62 Binz et al. 1990	
LH ₄₂₃ /A	153/7	23LH423/A (Q2E, N26K, A27Y)		Bol Ilin son	Bol Ilinu al.	
Ĭ,	Ŧ	بار پت	H. 32	ve otu mp:	ve ootu	
	23[Q	ر (ک	Vati 7. <i>b</i>	Vati C. b Sinz	
				-201	200	

= REGIONS OF NON-IDENTITY WITH THE NATIVE SEQUENCES.

-16.4

1321/441 TCA TTA GAT AAA GGA TAC AAT AAG agc gct gat ggg GCA TTA AAT GAT TTA TGT ATC AAA S L D K G Y N K S A D G A L N D L C I K Eco47 III

1321/441 TCA TTA GAT AAA GGA TAC AAT AAG atc gaa ggt cgt tgc gat ggg GCA TTA AAT GAT S L D K G Y N K I E G R C D G A L N D Factor Xa protease motif

AGG R ACA T GAC D CAG ATG M 66A 6 CCT P GAG E TCT 161 C 900 8 AAG K AGG R CT 6 6TG V CGG R AGG R CA6 AGT S AGC S 2617/873 TTT ACT F F T 2677/893 GAT GCT D A 2737/913 GGC TCC G S S 2797/933 TGT GAT GAA E GT.G V TAT Y GCT A CGG R CTG L 999 9 GAG E ACA T TTC F ک م AAG K GCT A 767 AGA R 999 AAG K GCC A 767 C CAA O 76C C AAC N GAG E T F GAT D GAT D رط د ط ACG T TAT Y GTG V 2587/863 TAC GTA (Y V L 2647/883 CCG GAG P E 2707/903 GGC TTT G F G F 2767/923 GGT ATC G I

-16.7

ACA CAT GAT D AAG K GAA E 2617/873 TTT ACT (F T 2677/893 TAT CAA T L L ATA I T L CAA O AGA R CAA O AGA R AAA K GAT D GAT 2587/863 TAC GTA (Y V [2647/883 TCT AAA (S K

F16.8

X AA CGA R GAA E CAA AAA K OCC A CAA OCA A GCA CTG L AAC N GAA E AAC N GCT ე<u>ე</u>ე A A A 77.C F TTA L GAA E GAA E CGA R TCA S AAA K GAA E E_ AAC N CAA AAG K AAC N AAC AAC N TTC F GAA E TTA L ATT I GAC D TTA L GCT A AAA K GAA E GTA V TAT Y AAC N AAC N AGC S AAC N ACA TTA L GAC ا اردو ا TTA L AGC S TT L GCG A GAG E AAA K CAG TTA GAC TAT Y TTA L TTA L TAT > GCT A ATC I GAT D AGA R GAT D GAG E K K CAG F A Q CA GCG A CAA TTA L A GCT N N TAT Y CTA L A GA AAT N AAC N ATC I TTC F AGT S GAT D 2587/863
TAC GTA GAT A
Y V D N
2647/883
TCC CCG GGT G
S P G A
2707/903
GAA CAA CAA A
E Q Q N
2767/923
AAC GCC TTC A
A F I Q
A K K L
2827/963
CAA AAC GCG I
Q N A F I
Q N A F
TTC ATC CAA A
F I Q S
3007/1003
AAG CTA AAT G
K L N D

F1G. 9

8/11 LH₄₂₃/A

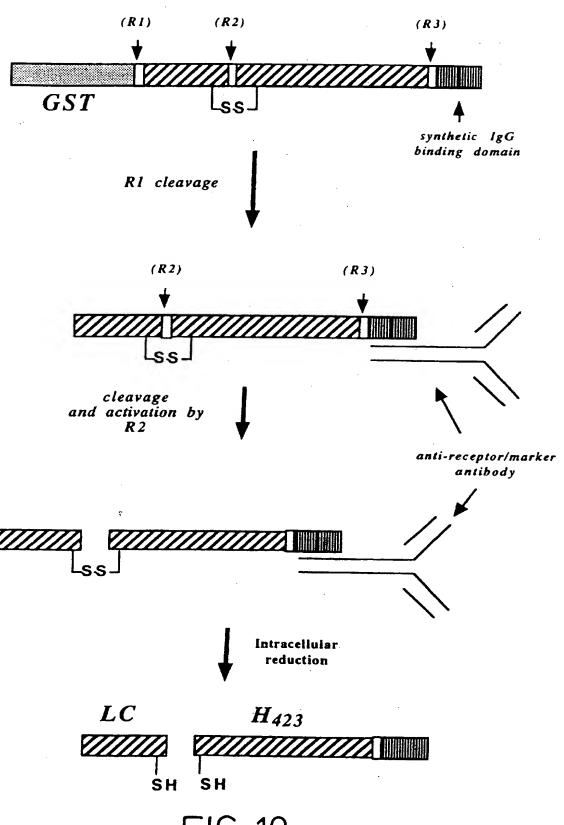
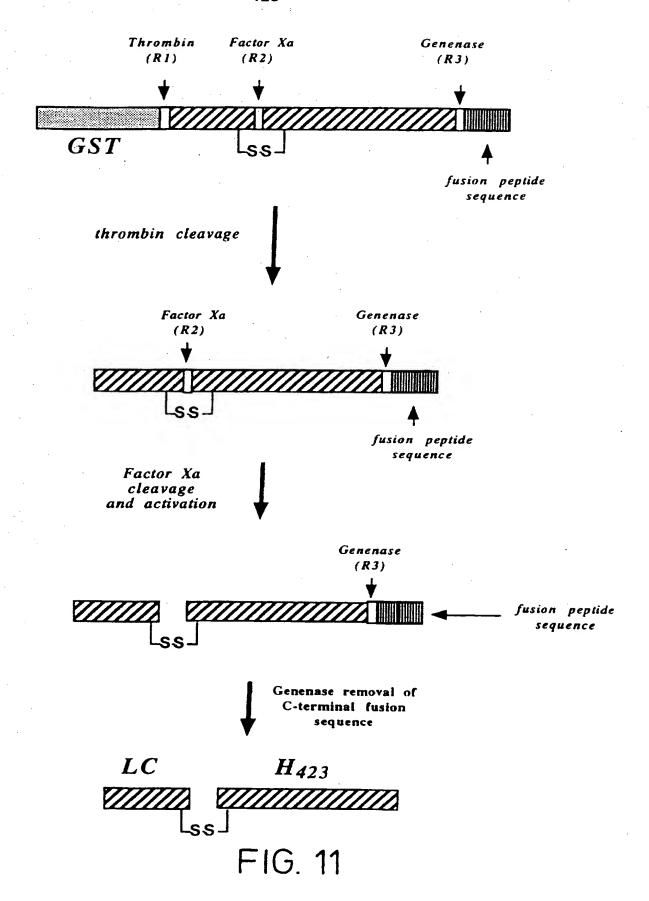


FIG. 10

$LH_{423}/A^{9/11}$



10/11

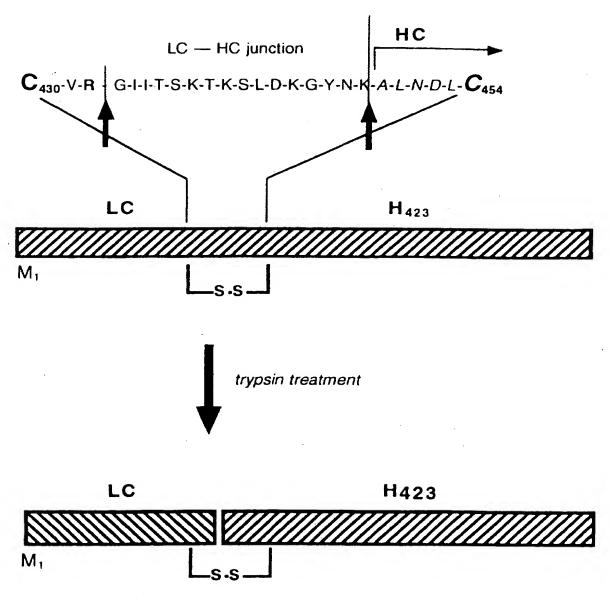
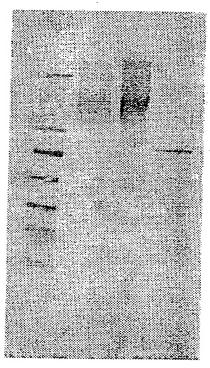


FIG. 12

Panel A.
1 2 3 4



Panel B. 1 2 3 4

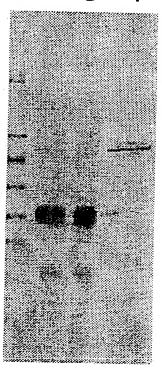


FIG. 13

INTERNATIONAL SEARCH REPORT

Inten onal Application No.

,		_		PCT/GB 9	7/02273
A. CLASS IPC 6	SFICATION OF SUBJECT MATTER C12N15/31 C12N1/21 C12N A61K39/08	P21/02	C07K14/	33 A61	K38/16
	to International Patent Classification (IPC) or to both national cl	lassification and	IPC		
	SEARCHED				
IPC 6	ocumentation searched (classification system followed by class C12N C12P A61K				
Documents	tion searched other than minimum documentation to the extent	that such docu	ments are includ	ed in the fields se	erched
Electronic	data base consulted during the international search (name of d	ata base and, v	rhere practical, s	earch terms used)
			·		
C. DOCUM	ENTS CONSIDERED TO BE RELEVANT				
Category *	Citation of document, with indication, where appropriate, of the	he relevant pasi	ages		Relevant to claim No.
X	WO 96 12802 A (OPHIDIAN PHARM; WILLIAMS JAMES A (US); PADHY (US); KI) 2 May 1996 see the whole document		V .		1-52
x	KURAZONO H ET AL: "Minimal es *domains* specifying toxicity *light* *chains* of tetanus to botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 P14721-9, UNITED STATES, XP002 see table II	of the oxin and (21)			1-52
		-/			
		,			
<u> </u>	or documents are listed in the continuation of box C.	X	atent family mon	nbers are fisted in	annex.
'A' document consider E' earlier do filing dat L' document which is citation of	t defining the general state of the art which is not red to be of particular relevance current but published on or after the international set which may throw doubts on priority claim(s) or cited to establish the publication date of another per or other special reason (as specified) t referring to an oral disclosure, use, exhibition or lans.	or pri cited inven "X" docum cann- invok "Y" docum cann- docum ment	ational filing date to application but try underlying the timed invention a considered to timent is taken alone timed invention nitive step when the to a person skilled		
r document later that	published prior to the international filing date but n the priority date claimed	in the docum**		ne same patent fa	mily

Date of mailing of the international search report

Hillenbrand, G

Authorized officer

3 0. 01. 98

Ruropean Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rüswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016 Form PCT/ISA/210 (second sheet) (July 1992)

Name and mailing address of the ISA

Date of the actual completion of the international search

9 December 1997

INTERNATIONAL SEARCH REPORT

Interconnal Application No PCT/GB 97/02273

C.(Continua	tion) DOCUMENTS CONSIDERED TO BE RELEVANT	
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938 see the whole document	1
A	BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348 see the whole document	1,26,35
	4.	
	·	
		.0]
	·	
·		

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inter. .onal Application No
PCT/GB 97/02273

Patent document cited in search report	Publication date		Patent family member(s)		Publication date
WO 9612802 A	02-05-96	US	5601823	Α.	11-02-97
		US	5196193	A	23-03-93
		AU	3968395	Α	15-05-96
	,	EP	0796326	Α	24-09-97
		FI	971732	Α	23-06-97
		NO	971868	A	24-06-97
		PL		Α	15-09-97
		ZA		Α	15-05-96
**		AU		Α	04-07-94
		CA		Α	23-06-94
		EP	0671902		20-09-95
	and the state of the		9413264	Α	23-06-94
		US		Α	14-11-95
		US	5599539		04-02-97
		AU	638786		08-07-93
		AU	6895191		31-05-91
		EP	0498854		19-08-92
		WO	9106306		16-05-91
		US	5443976		22-08-95
		US	5340923	Α	23-08-94

THIS PAGE BLANK (USPTO)

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:	
☐ BLACK BORDERS	
☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES	
☐ FADED TEXT OR DRAWING	
☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING	
☐ SKEWED/SLANTED IMAGES	
☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS	
☐ GRAY SCALE DOCUMENTS	
☐ LINES OR MARKS ON ORIGINAL DOCUMENT	
☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY	
Потигр.	

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)